ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Computer and System Sciences 69 (2004) 448–484

JOURNAL OF COMPUTER AND SYSTEM SCIENCES

www.elsevier.com/locate/jcss

Classical complexity and quantum entanglement

Leonid Gurvits

Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Received 14 September 2003; received in revised form 20 May 2004

Abstract

Generalizing a decision problem for bipartite perfect matching, Edmonds (J. Res. Natl. Bur. Standards 718(4) (1967) 242) introduced the problem (now known as the Edmonds Problem) of deciding if a given linear subspace of $M(N)$ contains a non-singular matrix, where $M(N)$ stands for the linear space of complex $N \times N$ matrices. This problem led to many fundamental developments in matroid theory, etc.

Classical matching theory can be defined in terms of matrices with non-negative entries. The notion of Positive operator, central in Quantum Theory, is a natural generalization of matrices with non-negative entries. (Here operator refers to maps from matrices to matrices.) First, we reformulate the Edmonds Problem in terms of completely positive operators, or equivalently, in terms of bipartite density matrices. It turns out that one of the most important cases when Edmonds' problem can be solved in polynomial deterministic time, i.e. an intersection of two geometric matroids, corresponds to unentangled (aka separable) bipartite density matrices. We introduce a very general class (or promise) of linear subspaces of $M(N)$ on which there exists a polynomial deterministic time algorithm to solve Edmonds' problem. The algorithm is a thoroughgoing generalization of algorithms in Linial, Samorodnitsky and Wigderson, Proceedings of the 30th ACM Symposium on Theory of Computing, ACM, New York, 1998; Gurvits and Yianilos, and its analysis benefits from an operator analog of permanents, so-called Quantum Permanents.

Finally, we prove that the weak membership problem for the convex set of separable normalized bipartite density matrices is NP-HARD.

Keywords: Entanglement; Complexity; Determinant

1. Introduction and main definitions

Let $M(N)$ be the linear space of $N \times N$ complex matrices. The following fundamental problem has been posed by Edmonds [10]:

E-mail address: gurvits@lanl.gov.

0022-0000/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcss.2004.06.003

Problem 1.1. Given a linear subspace $V \subset M(N)$ to decide if there exists a non-singular matrix $A \in V$.

We will assume throughout the paper that the subspace V is presented as a finite spanning k-tuple of rational matrices $S(V) = \{A_1, \ldots, A_k\}$ ($k \leqslant N^2$), i.e. the linear space generated by them is equal to V. As usual, the complexity parameter of the input, $< S(V) >$, is equal to (N + "number of bits of entries of matrices A_i, $1 \leqslant i \leqslant k$").

Edmonds' problem is equivalent to checking if the following determinantal polynomial:

$$P_{\mathbf{A}}(x_1, \ldots, x_k) = \det\left(\sum_{1 \leqslant i \leqslant k} x_i A_i\right)$$

is not identically equal to zero. The value of the determinantal polynomial at a particular point can be evaluated efficiently, hence randomized poly-time algorithms, based on Schwartz's lemma or its recent improvements, are readily available (notice that our problem is defined over an infinite field with infinite characteristic). But for general linear subspaces of $M(N)$, i.e. without an extra assumption (promise), poly-time deterministic algorithms are not known. Moreover, in light of the recent breakthrough paper [24] and Valiant's result [31] on universality of symbolic determinants, the deterministic complexity of Edmonds' problem has become fundamentally important in theoretical computer science.

Like any other homogeneous polynomial, $P_{\mathbf{A}}(x_1, \ldots, x_k)$ is a weighted sum of monomials of degree N, i.e.

$$P_{\mathbf{A}}(x_1, \ldots, x_k) = \sum_{(r_1, \ldots, r_k) \in I_{k,N}} a_{r_1, \ldots, r_k} x_1^{r_1} x_2^{r_2} \ldots x_k^{r_k}, \tag{1}$$

where $I_{k,N}$ stands for a set of vectors $r = (r_1, \ldots, r_k)$ with non-negative integer components and $\sum_{1 \leqslant i \leqslant k} r_i = N$. We will make substantial use of the following (Hilbert) norm of homogeneous polynomials, which we call the "G-norm":
$Q(x_1, \ldots, x_k) = \sum_{(r_1, \ldots, r_k) \in I_{k,N}} b_{r_1, \ldots, r_k} x_1^{r_1} x_2^{r_2} \ldots x_k^{r_k}$:

$$\|Q\|_G^2 =: \sum_{(r_1, \ldots, r_k) \in I_{k,N}} |b_{r_1, \ldots, r_k}|^2 r_1! r_2! \ldots r_k!. \tag{2}$$

It is easy to show that the determinantal polynomial $P_{\mathbf{A}}(x_1, \ldots, x_k) \equiv 0$ iff $P_{\mathbf{A}}(r_1, \ldots, r_k) = 0$ for any $|I_{k,N}| = \frac{(N+k-1)!}{N!(k-1)!}$ distinct points, in particular if it is zero for all $(r_1, \ldots, r_k) \in I_{k,N}$, which amounts to $\frac{(N+k-1)!}{N!(k-1)!}$ computations of determinants. We will show that $\|P_{\mathbf{A}}\|_G^2$ can be evaluated in $O(2^N N!)$ computations of determinants. If $k > \frac{2}{e^2} N^2$ then our approach is exponentially faster than computing $|I_{k,N}|$ determinants. More importantly, $\|P_{\mathbf{A}}\|_G^2$ serves as a natural tool to analyze our main algorithm.

The algorithm to solve Edmonds' problem, which we introduce and analyze later in the paper, is a rather thoroughgoing generalization of the recent algorithms [25,23] for deciding the existence of perfect matchings. They are based on so-called Sinkhorn's iterative scaling. The algorithm in [23] is a greedy version of Sinkhorn's scaling and has been analyzed using KLD-divergence; the algorithm in [25] is a standard Sinkhorn's scaling and a "potential" used for its analysis is the permanent. Our analysis is a sort of combination of techniques from [25,23]. Most importantly, $\|P_{\mathbf{A}}\|_G^2$ can be viewed as a generalization of the permanent.

The organization of this paper proceeds as follows. In Section 2 we will recall basic notions from Quantum Information Theory such as bipartite density matrix, positive and completely positive operator, separability and entanglement. After that we will rephrase Edmonds' problem using those notions and reformulate the famous Edmonds–Rado theorem on the rank of intersection of two geometric matroids in terms of the rank non-decreasing property of the corresponding (separable) completely positive operator. We will end Section 2 by introducing a property, called the Edmonds–Rado property, of linear subspaces of $M(N)$ which allows a poly-time deterministic algorithm to solve Edmonds' problem and will explain how is this property is related to quantum entanglement (see Theorem 2.7). In Section 3 we will express the G-norm of a determinantal polynomial $P_{\mathbf{A}}(x_1, \ldots, x_k)$ in terms of the associated bipartite density matrix, and we will prove various inequalities and properties of the G-norm which will be needed later on for the analysis of the main algorithm. In Section 4 we will introduce and analyze the main algorithm of the paper, operator Sinkhorn scaling. In Section 5 we will apply this algorithm to solve Edmonds' problem for linear subspaces of $M(N)$ having the Edmonds–Rado property. In Section 6 we will prove NP-HARDNESS of the weak membership problem for the compact convex set of separable normalized density matrices. Finally, in the Conclusion section we will pose several open problems and directions for future research.

The main algorithm of this paper is the third generation in a series of "scaling" procedures applied to computer science problems. These began with [25,23] (applied to bipartite perfect matchings and an approximation of the permanent), followed by Gurvits and Samorodnitsky [21,22] (applied to an approximation of the mixed discriminant and mixed volume). Here it is used to solve a very non-trivial, important and seemingly different problem.

2. Bipartite density matrices, completely positive operators and Edmonds problem

Definition 2.1. A positive semidefinite matrix $\rho_{A,B} : C^N \otimes C^N \to C^N \otimes C^N$ is called a bipartite unnormalized density matrix (BUDM). If $tr(\rho_{A,B}) = 1$ then this $\rho_{A,B}$ is called a bipartite density matrix. It is convenient to represent a bipartite $\rho_{A,B} = \rho(i_1, i_2, j_1, j_2)$ as the following block matrix:

$$\rho_{A,B} = \begin{pmatrix} A_{1,1} & A_{1,2} & \ldots & A_{1,N} \\ A_{2,1} & A_{2,2} & \ldots & A_{2,N} \\ \ldots & \ldots & \ldots & \ldots \\ A_{N,1} & A_{N,2} & \ldots & A_{N,N} \end{pmatrix}, \tag{3}$$

where $A_{i_1,j_1} =: \{\rho(i_1, i_2, j_1, j_2) : 1 \leqslant i_2, j_2 \leqslant N\}$, $1 \leqslant i_1, j_1 \leqslant N$. We interpret the "which-block" indices i_1, j_1 as referring to a system "A", and the indices i_2, j_2 of elements of the matrices that form the blocks, as referring to a system "B".

A BUDM ρ is called *separable* if there exist K-tuples $X := [x_1, \ldots, x_K]$ and $Y := [y_1, \ldots, y_K]$ of vectors in C^N such that

$$\rho = \rho_{(X,Y)} =: \sum_{1 \leqslant i \leqslant K} x_i x_i^\dagger \otimes y_i y_i^\dagger, \tag{4}$$

and *entangled* otherwise. (The RHS defines the notation $\rho_{(X,Y)}$.) In quantum information theory, separability (usually applied to *normalized* density matrices, i.e. BUDM whose trace is unity) is the formal

definition of the notion, "not entangled": the state can be written as a convex combination of pure quantum states that, since they are tensor products, show no correlation between A and B.

If the vectors $x_i, y_i; 1 \leqslant i \leqslant K$ in (4) are real then ρ is called *real separable*.

The quantum marginals are defined as $\rho_B = \sum_{1 \leqslant i \leqslant N} A_{i,i}$ and $\rho_A(i, j) = tr(A_{i,j}); 1 \leqslant i, j \leqslant N$. (In quantum information theory, these are sometimes written $\rho_B = tr_A \rho_{A,B}, \rho_A = tr_B \rho_{A,B}$.)

Next we define the BUDM $\rho_{\mathbf{A}}$ associated with the k-tuple $\mathbf{A} = (A_1, \ldots, A_k)$:

$$\rho_{\mathbf{A}}(i_1, i_2, j_1, j_2) =: \sum_{1 \leqslant l \leqslant k} A_l(i_2, i_1)\overline{A_l(j_2, j_1)}, \tag{5}$$

where for a complex number $z = x + iy$ its conjugate $\bar{z} = x - iy$. Rewriting expression (5) in terms of blocks of $\rho_{\mathbf{A}}$ as in (3), we get that

$$A_{i,j} = \sum_{1 \leqslant l \leqslant k} A_l e_i e_j^{\dagger} A_l^{\dagger}, \ 1 \leqslant i, j \leqslant N.$$

Remark 2.2. There is a natural (column by column) correspondence between $M(N)$ and $C^{N^2} \cong C^N \otimes C^N$. It works as follows:

$$A \equiv \{A(i, j), 1 \leqslant i, j \leqslant N\} \in M(N) \Leftrightarrow$$

$$v_A = (A(1, 1), \ldots, A(1, N); \ldots\ldots; A(1, N), \ldots, A(N, N))^T \in C^{N^2}.$$

Interpreting this correspondence in the language of quantum physics, one can view a matrix A as an (unnormalized) *pure* quantum state ("wavefunction") of a bipartite system, by interpreting its matrix elements $A(i, j)$ as the components of the state vector (element of $C^N \otimes C^N$) in a product basis $e_i \otimes e_j$ for $C^N \otimes C^N$. Then we may interpret a k-tuple $\mathbf{A} = (A_1, \ldots, A_k)$ of complex matrices as a k-tuple of unnormalized bipartite "wave functions," and the BUDM $\rho_{\mathbf{A}}$ as the corresponding mixed bipartite state formed as the sum of the (not-necessarily-normalized) pure states $v_A v_A^{\dagger}$. $Im(\rho_{\mathbf{A}})$ is, of course, the span of the vectors v_A.

We will call a BUDM ρ weakly separable if there exists a separable $\rho'_{(X,Y)}$ with the same image as ρ: $Im(\rho) = Im(\rho'_{(X,Y)})$. (Recall that in this finite dimensional case $Im(\rho)$ is the linear subspace formed by all linear combinations of columns of matrix ρ.)

A linear operator $T : M(N) \to M(N)$ is called positive if $T(X) \succeq 0$ for all $X \succeq 0$, and strictly positive if $T(X) \succeq \alpha\, tr(X) I$ for all $X \succeq 0$ and some $\alpha > 0$. A positive operator T is called completely positive if there are $A_i \in M(N)$ such that

$$T(X) = \sum_{1 \leqslant i \leqslant N^2} A_i X A_i^{\dagger}; \forall X \in M(N). \tag{6}$$

Choi's representation of a linear operator $T : M(N) \to M(N)$ is a block matrix $CH(T)_{i,j} =: T(e_i e_j^{\dagger})$. The dual to T with respect to the inner product $< X, Y >= tr(XY^{\dagger})$ is denoted as T^*. (Notice that if T

is completely positive and

$$T(X) = \sum_{1 \leqslant i \leqslant N^2} A_i X A_i^{\dagger};\ A_i, X \in M(N),$$

then $T^*(X) = \sum_{1 \leqslant i \leqslant N^2} A_i^{\dagger} X A_i$.)

A very useful and easy result of Choi [9] states that T is completely positive iff $CH(T)$ is a BUDM (i.e., positive semidefinite).

Using this natural (linear) correspondence between completely positive operators and BUDM, we will freely "transfer" properties of BUDM to completely positive operators. For example, a linear operator T is called separable iff $CH(T)$ is separable, i.e. if there exist K-tuples X and Y of vectors $x_i, y_i \in C^N$ such that

$$T(Z) = T_{(X,Y)}(Z) := \sum_{1 \leqslant i \leqslant K} x_i y_i^{\dagger} Z y_i x_i^{\dagger}. \tag{7}$$

Notice that $CH(T_{(X,Y)}) = \rho_{(\bar{Y},X)}$ and $T^*_{(X,Y)} = T_{(Y,X)}$. (The components of the vector $\bar{y}$ are the complex conjugates of corresponding components of y).

In light of definition (2.1), we will represent a linear subspace $V \subset M(N) \cong C^N \otimes C^N$ in Edmonds Problem as the image of the BUDM ρ. And as the complexity measure we will use the number of bits of (rational) entries of ρ plus the dimension N.

Definition 2.3. A positive linear operator $T : M(N) \to M(N)$ is called rank non-decreasing iff

$$Rank(T(X)) \geqslant Rank(X) \quad \text{if } X \succeq 0, \tag{8}$$

and is called indecomposable iff

$$Rank(T(X)) > Rank(X) \quad \text{if } X \succeq 0 \text{ and } 1 \leqslant Rank(X) < N. \tag{9}$$

A positive linear operator $T : M(N) \to M(N)$ is called doubly stochastic iff $T(I) = I$ and $T^*(I) = I$; it is called ε - doubly stochastic iff $DS(T) =: tr((T(I) - I)^2) + tr((T^*(I) - I)^2) \leqslant \varepsilon^2$.

The next Proposition 2.4 is a slight generalization of the corresponding result in [25].

Proposition 2.4. *Doubly stochastic operators are rank non-decreasing. Suppose that linear positive operator $T : M(N) \to M(N)$.*
If either $T(I) = I$ or $T^*(I) = I$ and $DS(T) \leqslant N^{-1}$ then T is rank non-decreasing. If $DS(T) \leqslant (2N+1)^{-1}$ then T is rank non-decreasing.

Proof. To prove the first, "N^{-1}", inequality we assume wlog that $T(I) = I$ and $T^*(I) = I + \Delta$, where Δ is a hermitian matrix and $tr(\Delta^2) \leqslant N^{-1}$. Let $U = (u_1, \ldots, u_N)$ be an orthonormal basis in C^N. Then

by linearity

$$\sum_{1\leqslant i\leqslant N} T(u_i u_i^\dagger) = T(I) = I.$$

Also

$$tr(T(u_i u_i^\dagger)) = tr(T(u_i u_i^\dagger)I) = tr(u_i u_i^\dagger T^*(I)) = 1+\delta_i, \quad 1\leqslant i\leqslant N,$$

where $\delta_i = tr(u_i u_i^\dagger \Delta)$. Clearly,

$$\sum_{1\leqslant i\leqslant N} |\delta_i|^2 \leqslant \sum_{1\leqslant i,j\leqslant N} |tr(u_i u_j^\dagger \Delta)|^2 = tr(\Delta^2) \leqslant N^{-1}.$$

Suppose that the positive operator T is not rank non-decreasing. That is there exists an orthonormal basis $U = (u_1, \ldots, u_N)$ such that for some $1\leqslant K\leqslant N-1$ the following rank inequality holds:

$$Rank\left(\sum_{1\leqslant i\leqslant K} T(u_i u_i^\dagger)\right) < K.$$

(This K is strictly less than N for $\sum_{1\leqslant i\leqslant N} T(u_i u_i^\dagger) = I$.)

Denote $A_i =: T(u_i u_i^\dagger), 1\leqslant i\leqslant N$. Since T is a positive operator $A_i \succeq 0, 1\leqslant i\leqslant N$. Therefore the matrix $H = \sum_{1\leqslant i\leqslant K} T(u_i u_i^\dagger)$ is positive semidefinite and $I \succeq H$. As $Rank(H)\leqslant K-1$ we get that $tr(H)\leqslant K-1$. On the other hand,

$$tr(H) = \sum_{1\leqslant i\leqslant K} tr(T(u_i u_i^\dagger)) = K + \sum_{1\leqslant i\leqslant k} \delta_i.$$

But $\sum_{1\leqslant i\leqslant n} |\delta_i|^2 \leqslant N^{-1}$. Therefore, the Cauchy–Schwarz inequality implies that

$$\sum_{1\leqslant i\leqslant k} |\delta_i| \leqslant \sqrt{\frac{K}{n}} < 1.$$

The last inequality contradicts to the inequality $tr(H)\leqslant K-1$. We got a desired contradiction.
The second, "$(2N+1)^{-1}$", inequality is proved using a similar application of the Cauchy–Schwarz inequality and left to the reader. □

Let us consider a completely positive operator $T_{\mathbf{A}} : M(N) \to M(N), T(X) = \sum_{1\leqslant i\leqslant k} A_i X A_i^\dagger$, and let $L(A_1, A_2, \ldots, A_K)$ be the linear subspace of $M(N)$ generated by matrices $\{A_i, 1\leqslant i\leqslant k\}$. It is easy to see that if $\hat{A} \in L(A_1, A_2, \ldots, A_k)$ then $\hat{A}(Im(X)) \subset Im(T(X))$ for all $X \succeq 0$. Therefore, if $L(A_1, A_2, \ldots, A_k)$ contains a non-singular matrix then the operator T is rank non-decreasing.

This simple observation suggested the following property of linear subspaces of $M(N)$.

Definition 2.5. A linear subspace $V = L(A_1, A_2, \ldots, A_k)$ has the *Edmonds–Rado Property* (*ERP*) if the existence of a non-singular matrix in V is equivalent to the fact that the associated completely positive operator $T_{\mathbf{A}}$ is rank non-decreasing.

In other words, a linear subspace $V \subset M(N)$ has the ERP if the fact that all matrices in V are singular is equivalent to the existence of two linear subspaces $X, Y \subset C^N$ such that $dim(Y) < dim(X)$ and $A(X) \subset Y$ for all matrices $A \in V$.

The main "constructive" result of this paper is that for linear subspaces of $M(N)$ having the ERP there is a deterministic poly-time algorithm to solve Edmonds' problem. In the rest of this section we will explain why we chose to call this property Edmonds–Rado, will describe a rather wide class of linear subspaces with the ERP and will give an example of a subspace without it.

2.1. Examples of linear subspaces of $M(N)$ having Edmonds–Rado Property

Let us first list some obvious but useful facts about the Edmonds–Rado property.

F1. Suppose that $V = L(A_1, A_2, \ldots, A_k) \subset M(N)$ has the ERP and $C, D \in M(N)$ are two non-singular matrices. Then the linear subspace $V_{C,D} =: L(CA_1D, CA_2D, \ldots, CA_kD)$ also has the ERP.

F2. If $V = L(A_1, A_2, \ldots, A_k) \subset M(N)$ has the ERP then both $V^\dagger =: L(A_1^\dagger, A_2^\dagger, \ldots, A_k^\dagger)$ and $V^T = L(A_1^T, A_2^T, \ldots, A_k^T)$ have the ERP.

F3. Any linear subspace $V = L(A_1, A_2, \ldots, A_k) \subset M(N)$ with matrices $\{A_i, 1 \leqslant i \leqslant k\}$ being positive semidefinite has the ERP.

F4. Suppose that linear subspaces $V = L(A_1, A_2, \ldots, A_k) \subset M(N_1)$ and $W = L(B_1, B_2, \ldots, B_k) \subset M(N_2)$ both have the ERP. Define the following matrices $C_i \in M(N_1 + N_2), 1 \leqslant i \leqslant k$:

$$C_i = \begin{pmatrix} A_i & D_i \\ 0 & B_i \end{pmatrix}.$$

Then the linear subspace $L(C_1, C_2, \ldots, C_k) \subset M(N_1 + N_2)$ also has the ERP.

A particular case of this fact is that any linear subspace of $M(N)$ which has a basis consisting of upper (lower) triangular matrices has the ERP.

F5. Any two-dimensional subspace of $M(N)$ has the ERP. In fact, for any two (but not three) square matrices $A, B \in M(N)$ there exist two non-singular matrices C, D such both CAD and are CBD upper (lower) triangular.

The next theorem gives the most interesting example which motivated the name "Edmonds–Rado Property". Let us first recall one of the most fundamental results in matroids theory, i.e. the Edmonds–Rado characterization of the rank of the intersection of two geometric matroids. A matroid is finite set (the "ground set") together with a set of subsets of that set satisfying properties abstracted from those of the set of all linearly independent subsets of a finite set of vectors in a linear space. A *geometric matroid* over C^N can be specified as a finite list of vectors $x_1, \ldots, x_K$ in C^N; this can be viewed as determining a matroid over the ground set $\{1, \ldots, K\}$, with the distinguished subsets being the subsets of $\{1, \ldots, K\}$ that correspond to linearly independent sets of vectors.

Definition 2.6. Let $X = (x_1, \ldots, x_K), Y = (y_1, \ldots, y_K)$ be two finite subsets of C^N, viewed as two geometric matroids on the ground set $\{1, 2, \ldots, K\}$. Their intersection $MI(X, Y) = \{(x_i, y_i), 1 \leqslant i \leqslant K\}$ is the set of distinct pairs of non-zero vectors (x_i, y_i). The rank of $MI(X, Y)$, denoted by $Rank(MI(X, Y))$ is the largest integer m such that there exist $1 \leqslant i_1 < \cdots < i_m \leqslant K$ with both sets $\{x_{i_1}, \ldots, x_{i_m}\}$ and $\{y_{i_1}, \ldots, y_{i_m}\}$ being linearly independent.

The Edmonds–Rado theorem (Corollary (7.5.17) in [14]) states (in the much more general situation of the intersection of any two matroids with a common ground set) that

$$Rank(MI(X,Y)) = \min_{S\subset\{1,2,\ldots,K\}} dim\, L(x_i;\, i\in S) + dim\, L(y_j;\, j\in \bar{S}). \tag{10}$$

(Note that $Rank(MI(X,Y))$ is the maximum rank achieved in the linear subspace $L(x_1y_1^\dagger,\ldots,x_Ky_K^\dagger)$ and that $Rank(MI(X,Y))=N$ iff $L(x_1y_1^\dagger,\ldots,x_Ky_K^\dagger)$ contains a non-singular matrix.)

Theorem 2.7. *Suppose that $T: M(N)\to M(N)$, $T(X)=\sum_{1\leqslant j\leqslant l} A_iXA_i^\dagger$, is a completely positive weakly separable operator, i.e. there exists a family of rank one matrices $\{x_1y_1^\dagger,\ldots,x_ly_l^\dagger\}\subset M(N)$ such that $L(A_1,\ldots,A_l)=L(x_1y_1^\dagger,\ldots,x_ly_l^\dagger)$. Then the following conditions are equivalent*:

Condition 1. *T is rank non-decreasing.*

Condition 2. *The rank of the intersection of the two geometric matroids $MI(X,Y)$ is equal to N.*

Condition 3. *There exists a non-singular matrix A such that for all $Z\succeq 0$, $Im(AZA^\dagger)\subset Im(T(Z))$.*

Condition 4. *There exists a non-singular matrix A such that the operator T' defined by $T'(Z)=T(Z)-AZA^\dagger$ is completely positive.*

Proof. ($2\Longrightarrow 1$): Suppose that the rank of $MI(X,Y)$ is equal to N. Then

$$Rank(T(Z))=dim(L(x_i;\, i\in S)),\quad \text{where } S=:\{i: y_i^\dagger Z y_i\neq 0\}.$$

As $dim(L(y_j;\, j\in\bar{S}))\leqslant dim(Ker(Z))=N-Rank(Z)$ hence, from the Edmonds–Rado Theorem we get that $Rank(T(Z))\geqslant N-(N-Rank(Z))=Rank(Z)$.

($1\Longrightarrow 2$): Suppose that T is rank non-decreasing and for any $S\subset\{1,2,\ldots,l\}$ consider an orthogonal projector $P\succeq 0$ on $L(y_j;\, j\in\bar{S})^\perp$. Then

$$dim(L(x_i: i\in S)) \geqslant Rank(T(P))\geqslant Rank(P)=N-dim(L(y_j;\, j\in\bar{S})).$$

It follows from the Edmonds–Rado Theorem that the rank of $MI(X,Y)$ is equal to N. All other "equivalences" follow now directly. □

Remark 2.8. Theorem 2.7 makes the Edmonds–Rado theorem sound like Hall's theorem on bipartite perfect matchings. Indeed, consider a weighted incidence matrix A_Γ of a bipartite graph Γ, i.e. $A_\Gamma(i,j)>0$ if i from the first part is adjacent to j from the second part and equal to zero otherwise. Then Hall's theorem can be immediately reformulated as follows:

A perfect matching, which is just a permutation in this bipartite case, exists iff $|A_\Gamma x|_+\geqslant|x|_+$ for any vector x with non-negative entries, where $|x|_+$ stands for the number of positive entries of a vector x.

All known algorithms (for instance, the linear programming algorithm presented in Section 7.5 (pp. 210–218)) of [14]) to compute the rank of the intersection of two geometric matroids require an

explicit knowledge of pairs of vectors (x_i, y_i), or, in other words, an explicit representation of the rank one basis $\{x_i y_i^\dagger, 1 \leqslant i \leqslant l\}$. The algorithm in this paper requires only a promise that such a rank one basis (not necessarily rational! [17]) does exist. That is it solves in polynomial deterministic time the following Hidden Matroids Intersection Problem (HMIP): *Given a linear subspace* $V \subset M(N)$, $V = L(A_1, \ldots, A_k)$ $(k \leqslant N^2)$, *where* A_i *are rational matrices, and a promise that* L *has a* (*hidden*) *basis consisting of rank one matrices* (*not necessarily rational!* [17]). *Check if the maximal rank achieved in* V *is equal* N.

It is unclear (to the author) what is the complexity of (HMIP) over finite fields.

Another example comes from [8]. Consider pairs of matrices $(A_i, B_i \in M(N);\ 1 \leqslant i \leqslant K)$. Let $V_i \subset M(N)$ be the linear subspace of all matrix solutions of the equation $XA_i = B_i X$. One of the problems solved in [8] is to decide if $W = V_1 \cap \cdots \cap V_K$ contains a non-singular matrix. It is not clear to the author whether the class of such linear subspaces W satisfies the ERP. But suppose that A_1 is similar to B_1 (V_1 contains a non-singular matrix) and, additionally, assume that $dim(Ker(A_1 - \lambda I)) = dim(Ker(B_1 - \lambda I)) \leqslant 1$ for all complex $\lambda \in C$ (i.e. there is just one Jordan block for each eigenvalue). It is not difficult to show that in this case there exist two non-singular matrices D, Q and upper triangular matrices $(U_1, \ldots, U_r)$ such that $V_1 = L(DU_1Q, \ldots, DU_rQ)$. It follows, using F1 and F4 from the beginning of this subsection, that V_1 as well as any of its linear subspaces has the ERP.

Example 2.9. Consider the following completely positive doubly stochastic operator $Sk_3 : M(3) \to M(3)$:

$$Sk_3(X) = \tfrac{1}{2}(A_{(1,2)} X A_{(1,2)}^\dagger + A_{(1,3)} X A_{(1,3)}^\dagger + A_{(2,3)} X A_{(2,3)}^\dagger). \tag{11}$$

Here $\{A_{(i,j)}, 1 \leqslant i < j \leqslant 3\}$ is a standard basis in the linear subspace $K(3) \subset M(3)$ consisting of all skew-symmetric matrices, i.e. $A_{(i,j)} =: e_i e_j^\dagger - e_j e_i^\dagger$ and $\{e_i, 1 \leqslant i \leqslant 3\}$ is a standard orthonormal basis in C^3.

It is clear that all 3×3 skew-symmetric matrices are singular. As Sk_3 is a completely positive doubly stochastic operator, and, thus, is rank non-decreasing, therefore $K(3) \subset M(3)$ is an example of a linear subspace not having ERP.

More "exotic" properties of this operator can be found in [17].

3. Quantum permanents and G-norms of determinantal polynomials

Consider a k-tuple of $N \times N$ complex matrices $\mathbf{A} = (A_1, \ldots, A_k)$. Our first goal here is to express the square of the G-norm of a determinantal polynomial $P_{\mathbf{A}}(x_1, \ldots, x_k)$ in terms of the associated bipartite density matrix (BUDM) $\rho_{\mathbf{A}}$, which is defined as in (5).

Consider an N-tuple of complex $N \times N$ matrices, $\mathbf{B} = (B_1, \ldots, B_N)$. Recall that the mixed discriminant $M(\mathbf{B}) = M(B_1, \ldots, B_N)$ is defined as follows:

$$M(B_1, \ldots B_N) = \frac{\partial^n}{\partial x_1 \ldots \partial x_N} det(x_1 B_1 + \cdots + x_N B_N). \tag{12}$$

Or equivalently

$$M(B_1, \dots B_N) = \sum_{\sigma,\tau \in S_N} (-1)^{sign(\sigma\tau)} \prod_{i=1}^{N} B_i(\sigma(i), \tau(i)), \tag{13}$$

where S_N is the symmetric group, i.e. the group of all permutations of the set $\{1, 2, \dots, N\}$. If matrices B_i, $1 \leqslant i \leqslant N$ are diagonal then their mixed discriminant is equal to the corresponding permanent [21].

Definition 3.1. Let us consider a block matrix $\rho = \{\rho(i_1, i_2, j_1, j_2)\}$ as in (3) (not necessarily positive semidefinite). Additionally to the blocks define also the following N^2-tuple of $N \times N$ matrices:

$$C_{j_1,j_2} =: \{\rho(i_1, i_2, j_1, j_2) : 1 \leqslant i_1, j_2 \leqslant N\}, 1 \leqslant j_1, j_2 \leqslant N.$$

We define the quantum permanent, $QP(\rho)$, by the following equivalent formulas:

$$QP(\rho) =: \sum_{\sigma \in S_N} (-1)^{sign(\sigma)} M(A_{1,\sigma(1)}, \dots, A_{N,\sigma(N)}), \tag{14}$$

$$QP(\rho) =: \sum_{\sigma \in S_N} (-1)^{sign(\sigma)} M(C_{1,\sigma(1)}, \dots, C_{N,\sigma(N)}), \tag{15}$$

$$QP(\rho) = \frac{1}{N!} \sum_{\tau_1,\tau_2,\tau_3,\tau_4 \in S_N} (-1)^{sign(\tau_1\tau_2\tau_3\tau_4)} \times \prod_{i=1}^{N} \rho(\tau_1(i), \tau_2(i), \tau_3(i), \tau_4(i)). \qquad \square \tag{16}$$

Straight from this definition, we get the following inner product formula for quantum permanents:

$$QP(\rho) = < \rho^{\otimes N} Z, Z >, \tag{17}$$

where $\rho^{\otimes N}$ stands for a tensor product of N copies of ρ, $< ., . >$ is a standard inner product and

$$Z(j_1^{(1)}, j_2^{(1)}; \dots; j_1^{(N)}, j_2^{(N)}) = \frac{1}{(N!)^{1/2}} (-1)^{sign(\tau_1\tau_2)}$$

if there exist two permutations $\tau_1, \tau_1 \in S_N$ such that $j_k^{(i)} = \tau_k(i) (1 \leqslant i \leqslant N, k = 1, 2)$, and $Z(j_1^{(1)}, j_2^{(1)}; \dots; j_1^{(N)}, j_2^{(N)}) = 0$ otherwise.

Remark 3.2. Notice that equality (17) implies that if $\rho_1 \succeq \rho_2 \succeq 0$ then $QP(\rho_1) \geqslant QP(\rho_2) \geqslant 0$. The standard norm of the N^{2N}-dimensional vector Z defined above is equal to 1. Thus, if ρ is a normalized bipartite density matrix then $QP(\rho)$ can be viewed as the probability of a particular outcome of some (von Neumann) measurement. Unfortunately, in this case $QP(\rho) \leqslant \frac{N!}{N^N}$.

Consider an arbitrary permutation $\sigma \in S_4$ and for a block matrix (or tensor) $\rho = \{\rho(i_1, i_2, i_3, i_4); 1 \leqslant i_1, i_2, i_3, i_4 \leqslant N\}$ define $\rho^\sigma = \{\rho(i_{\sigma(1)}, i_{\sigma(2)}, i_{\sigma(3)}, i_{\sigma(4)})\}$. It is easy to see that

$QP(\rho) = QP(\rho^{\sigma})$. Another simple but important fact about quantum permanents is the following identity:

$$QP((A_1 \otimes A_2)\rho(A_3 \otimes A_4)) = \det(A_1A_2A_3A_4)QP(\rho). \tag{18}$$

The author recently learned that the "quantum permanent" is a four-dimensional version of a 19th century generalization of the determinant by Ernst Pascal [28]. Theorem 3.8 can be easily generalized to the case of $2d$-dimensional Pascal's determinants, which will give unbiased (but not well concentrated) estimators of $2d$-dimensional Pascal's determinants in terms of the usual (i.e. two-dimensional) determinants.

The author clearly (and sympathetically) realizes that some readers might object to (or ridicule) the name "quantum permanent." The next example, hopefully, will explain possible motivations.

Example 3.3. Let us present a few cases when Quantum Permanents can be computed "exactly." They will also illustrate how universal this new notion is:

1. Let $\rho_{A,B}$ be a product state, i.e. $\rho_{A,B} = C \otimes D$. Then $QP(C \otimes D) = N!Det(C)Det(D)$.
2. Let $\rho_{A,B}$ be a pure state, i.e. there exists a matrix $(R = R(i, j) : 1 \leqslant i, j \leqslant N)$ such that $\rho_{A,B}(i_1, i_2, j_1, j_2) = R(i_1, i_2)\overline{R(j_1, j_2)}$. In this case $QP(\rho_{A,B}) = N!|Det(R)|^2$.
3. Define blocks of $\rho_{A,B}$ as $A_{i,j} = R(i, j)e_ie_j^{\dagger}$. Then $QP(\rho_{A,B}) = Per(R)$.
4. Consider a **separable** BUDM represented (nonuniquely) as

$$\rho = \rho_{(X,Y)} =: \sum_{1 \leqslant i \leqslant K} x_i x_i^{\dagger} \otimes y_i y_i^{\dagger},$$

Define the matroidal permanent as

$$MP_{(X,Y)} =: \sum_{1 \leqslant i_1 < i_2 < \cdots < i_N \leqslant K} det\left(\sum_{1 \leqslant k \leqslant N} x_{i_k} x_{i_k}^{\dagger}\right) det\left(\sum_{1 \leqslant k \leqslant N} y_{i_k} y_{i_k}^{\dagger}\right).$$

It follows from Theorem 3.8 that $MP_{(X,Y)} = QP(\rho_{(X,Y)})$.

The following propositions provide important upper bounds for quantum permanents of positive semidefinite matrices.

Proposition 3.4. *Suppose that $\rho_{A,B}$ is a* **BUDM**. *Then*

$$\max_{\sigma \in S_N} |M(A_{1,\sigma(1)}, \ldots, A_{N,\sigma(N)})| = M(A_{1,1}, \ldots, A_{N,N}) \tag{19}$$

Proof. For $\tau, \sigma \in S_N$ define a matrix

$$B_{\tau,\sigma} =: A_{\tau(1),\sigma(1)} \otimes A_{\tau(2),\sigma(2)} \otimes \cdots \otimes A_{\tau(N),\sigma(N)}$$

Since $\rho_{A,B}$ is positive semidefinite hence the block matrix $\{B_{\tau,\sigma} : \tau, \sigma \in S_N\}$ is also positive semidefinite. It is well known [1] and easy to prove that

$$M(A_1, \ldots, A_N) = tr((A_1 \otimes \cdots \otimes A_N)VV^\dagger),$$

where the N^N-dimensional vector $V = V(i_1, i_2, \ldots, i_n) : 1 \leqslant i_k \leqslant N, 1 \leqslant k \leqslant N$ is defined as follows: $V(i_1, i_2, \ldots, i_n) = \frac{1}{\sqrt{N!}}(-1)^{sign(\tau)}$ if there exists a permutation $\tau \in S_N$; and equal to zero otherwise. It follows that the following $N! \times N!$ matrix C

$$C_{\tau,\sigma} = tr(B_{\tau,\sigma}VV^\dagger) = M(A_{\tau(1),\sigma(1)}, A_{\tau(2),\sigma(2)}, \ldots, A_{\tau(N),\sigma(N)})$$

is also positive semidefinite. Thus

$$|C_{\tau,\sigma}| \leqslant (C_{\tau,\tau}C_{\sigma,\sigma})^{\frac{1}{2}} = M(A_{1,1}, \ldots, A_{N,N}).$$

Corollary 3.5. *If $\rho_{A,B}$ is a* BUDM *then*

$$QP(\rho_{A,B}) \leqslant N!M(A_{1,1}, \ldots, A_{N,N}) \leqslant N!Det(\rho_B). \tag{20}$$

(*The permanental part of Example 3.3 shows that $N!$ is the exact constant in both parts of* (20)*, i.e. if the blocks $A_{i,j} = e_i e_j^\dagger$, $1 \leqslant i, j \leqslant N$ then $QP(\rho_{A,B}) = N!$ and $M(A_{1,1}, \ldots, A_{N,N}) = Det(\rho_A) = 1$.*)

The next proposition follows from Hadamard's inequality, which states that if $X \succ 0$ is $N \times N$ matrix then $Det(X) \leqslant \prod_{i=1}^N X(i,i)$.

Proposition 3.6. *If $X \succ 0$ then the following inequality holds*:

$$Det\left(\sum_{i=1}^K x_i y_i^\dagger X y_i x_i^\dagger\right) \geqslant Det(X)MP_{(X,Y)}. \tag{21}$$

(*The quantity $MP_{(X,Y)}$ is defined in part 4 of Example 3.3. In this separable case $MP_{(X,Y)} = QP(\rho_{(X,Y)})$, where $\rho_{(X,Y)} = \sum_{1 \leqslant i \leqslant K} x_i x_i^\dagger \otimes y_i y_i^\dagger$.*)

Corollary 3.7. *Suppose that a separable BUDM $\rho_{A,B}$ is Choi's representation of the completely positive operator T. Then for all $X \succ 0$ the following inequality holds*:

$$Det(T(X)) \geqslant QP(\rho_{A,B})Det(X). \tag{22}$$

Since $\rho_B = T(I)$, hence $QP(\rho_{A,B}) \leqslant Det(\rho_B)$ in the separable case.

(Notice that Corollary 3.5 provides an example of an entangled BUDM which does not satisfy (22).)

Finally, in the next theorem we connect quantum permanents with G-norms of determinantal polynomials.

Theorem 3.8. 1. *Consider an arbitrary polynomial*

$$P(x_1, x_2, \ldots, x_k) = \sum_{(r_1,\ldots,r_k)\in F} a_{r_1,\ldots,r_k} x_1^{r_1} x_2^{r_2} \ldots x_k^{r_k}, \ |F| < \infty,$$

where F is some finite set of vectors with non-negative integer components and define its G-norm as follows

$$\|P\|_G^2 =: \sum_{(r_1,\ldots,r_k)\in F} |a_{r_1,\ldots,r_k}|^2 r_1! r_2! \ldots r_k!$$

Then the following identity holds:

$$\|P\|_G^2 = E_{\xi_1,\ldots,\xi_k}(|P(\xi_1, \ldots, \xi_k)|^2), \tag{23}$$

where $(\xi_1, \ldots, \xi_k)$ are independent identically distributed zero mean gaussian complex random variables and the covariance matrix of ξ_1, viewed as a 2×2 real matrix, is equal to $\frac{1}{2} I$.

2. *Consider a k-tuple of $N \times N$ complex matrices $\mathbf{A} = (A_1, \ldots, A_k)$ and the corresponding determinantal polynomial $P_{\mathbf{A}}(x_1, \ldots, x_k) =: det(\sum_{1\leqslant i\leqslant k} x_i A_i)$. Then the following identity holds*:

$$\|P_{\mathbf{A}}\|_G^2 = QP(\rho_{\mathbf{A}}). \tag{24}$$

Proof. The two proofs ("long" deterministic and "short" probabilistic) can be found in Appendices A and C. □

Remark 3.9. Theorem 3.8, more precisely the combination of its two parts, can be viewed as a generalization of the famous Wick formula [34]. It seems reasonable to predict that formula (24) might be of use in the combinatorics described in [34].

It is well known (see, for instance, [2]) that the mixed discriminant $M(A_1, \ldots, A_N)$ can be evaluated by computing 2^N determinants. Therefore the quantum permanent $QP(\rho)$ can be evaluated by computing $N!2^N$ determinants. Now, formula (24) suggests the following algorithm to compute $\|det(\sum_{1\leqslant i\leqslant k} x_i A_i)\|_G^2$: first, construct the associated bipartite density matrix $\rho_{\mathbf{A}}$, which will take $O(N^4 k)$ arithmetic operations (additions and multiplications); secondly, compute $QP(\rho_{\mathbf{A}})$.

The total cost of this procedure (number of arithmetic operations) is $Cost(N) = O(N!2^N N^3)$. On the other hand, just the number of monomials in $det(\sum_{1\leqslant i\leqslant k} x_i A_i)$ is equal to $|I_{k,N}| = \frac{(N+k-1)!}{N!(k-1)!}$. If $k - 1 = aN^2$ then $|I_{k,N}| \geqslant \frac{a^N N^{2N}}{N!}$. Thus,

$$\frac{|I_{k,N}|}{Cost(N)} \geqslant \frac{a^N N^{2N}}{O(N!^2 2^N N^3)} \geqslant \approx \frac{(ae^2)^N}{2^N N^3}.$$

We conclude that if $a > \frac{2}{e^2}$ our approach is exponentially faster than the "naive" one, i.e. than evaluating $det(\sum_{1\leqslant i\leqslant k} x_i A_i)$ at all vectors $(x_1, \ldots, x_k) \in I_{k,N}$.

Our approach provides an $O(N!2^N N^3)$-step deterministic algorithm to solve the general case of Edmonds' Problem.

4. Operator Sinkhorn's iterative scaling

Recall that for a square matrix $A = \{a_{ij} : 1 \leqslant i, j \leqslant N\}$ row scaling is defined as

$$R(A) = \left\{ \frac{a_{ij}}{\sum_j a_{ij}} \right\},$$

column scaling as $C(A) = \{\frac{a_{ij}}{\sum_i a_{ij}}\}$ assuming that all denominators are non-zero.

The iterative process $\ldots CRCR(A)$ is called *Sinkhorn's iterative scaling* (SI). There are two main, well known, properties of this iterative process, which we will generalize to positive operators.

Proposition 4.1. 1. *Suppose that* $A = \{a_{i,j} \geqslant 0 : 1 \leqslant i, \ j \leqslant N\}$. *Then* (*SI*) *converges iff A is matching, i.e., there exists a permutation* π *such that* $a_{i,\pi(i)} > 0$ $(1 \leqslant i \leqslant N)$ [7].

2. *If A is indecomposable, i.e., A has a doubly-stochastic pattern and is fully indecomposable in the usual sense, then* (*SI*) *converges exponentially fast. Also in this case there exist unique positive diagonal matrices* $D_1, D_2, \det(D_2) = 1$ *such that the matrix* $D_1^{-1} A D_2^{-1}$ *is doubly stochastic.* (*For this part see, for instance* [30].)

Definition 4.2. (Operator scaling). Consider a positive linear operator $T : M(N) \to M(N)$. Define a new positive operator, Operator scaling, $S_{C_1,C_2}(T)$ as

$$S_{C_1,C_2}(T)(X) =: C_1 T(C_2^\dagger X C_2) C_1^\dagger. \tag{25}$$

Assuming that both $T(I)$ and $T^*(I)$ are non-singular we define analogs of row and column scalings:

$$R(T) = S_{T(I)^{-\frac{1}{2}},I}(T), \ C(T) = S_{I,T^*(I)^{-\frac{1}{2}}}(T). \tag{26}$$

Operator Sinkhorn's iterative scaling (OSI) is the iterative process $\ldots CRCR(T)$.

Remark 4.3. Using Choi's representation of the operator T as in Definition 2.1, we can define analogs of operator scaling (which are exactly the so called local transformations in Quantum Information Theory) and OSI in terms of BUDM:

$$\begin{aligned} S_{C_1,C_2}(\rho_{A,B}) &= C_1 \otimes C_2(\rho_{A,B}) C_1^\dagger \otimes C_2^\dagger, \\ R(\rho_{A,B}) &= \rho_B^{-\frac{1}{2}} \otimes I(\rho_{A,B}) \rho_B^{-\frac{1}{2}} \otimes I, \\ C(\rho_{A,B}) &= I \otimes \rho_A^{-\frac{1}{2}}(\rho_{A,B}) I \otimes \rho_A^{-\frac{1}{2}}. \end{aligned} \tag{27}$$

The standard ("classical") *Sinkhorn's iterative scaling* is a particular case of Operator Sinkhorn's iterative scaling (OSI) when the initial Choi's representation of the operator T is a diagonal BUDM.

Let us introduce a class of locally scalable functionals LSF defined on a set of positive linear operators, i.e. functionals satisfying the following identity:

$$\varphi(S_{C_1,C_2}(T)) = det(C_1 C_1^\dagger) det(C_2 C_2^\dagger) \varphi(T). \tag{28}$$

We will call an LSF bounded if there exists a function f such that $|\varphi(T)| \leqslant f(tr(T(I))$. It is clear that bounded LSF are natural "potentials" for analyzing (OSI). Indeed, Let $T_n, T_0 = T$ be a trajectory of (OSI). T is a positive linear operator. Then $T_i(I) = I$ for odd i and $T^*_{2i}(I) = I, i \geqslant 1$. Thus if $\varphi(.)$ is LSF then

$$\varphi(T_{i+1}) = a(i)\varphi(T_i),\ a(i) = Det(T_i^*(I))^{-1} \text{ if } i \text{ is odd},$$
$$a(i) = Det(T_i(I))^{-1} \text{ if } i > 0 \text{ is even}. \tag{29}$$

As $tr(T_i(I)) = tr(T_i^*(I)) = N, i > 0$, thus by the arithmetic/geometric means inequality we have that $|\varphi(T_{i+1})| \geqslant |\varphi(T_i)|$ and if $\varphi(.)$ is bounded and $|\varphi(T)| \neq 0$ then $DS(T_n)$ converges to zero.

To prove a generalization of Statement 1 in Proposition 4.1 we need to "invent" a bounded LSF $\varphi(.)$ such that $\varphi(T) \neq 0$ iff the operator T is rank non-decreasing on positive matrices. We call such functionals "responsible for matching". It follows from (18) and (20) that $QP(CH(T))$ is a bounded LSF if T is a completely positive operator. (In fact $QP(CH(T))$ is a bounded LSF even if T is just positive.) Thus if $QP(CH(T)) \neq 0$ then $DS(T_n)$ converges to zero and, by Proposition 2.4, T is rank non-decreasing. On the other hand, in Example 2.8 $QP(CH(Sk_3)) = 0$ and Sk_3 is rank non-decreasing (even indecomposable), i.e. $QP(CH(T))$ in general is not responsible for matching even if T is a completely positive operator. (This is another "strangeness" of entangled operators. We wonder if it is possible to have a "nice," say polynomial with integer coefficients, responsible for matching LSF?)
We now introduce a responsible for matching bounded LSF, the capacity of a positive operator, which is continuous but non-differentiable.

Definition 4.4. For a positive operator $T : M(N) \to M(N)$, we define its capacity as

$$Cap(T) = \inf\{Det(T(X)) : X \succ 0, Det(X) = 1\}.$$

It is easy to see that $Cap(T)$ is LSF. Since $Cap(T) \leqslant Det(T(I)) \leqslant (\frac{tr(T(I))}{N})^N$, $Cap(T)$ is a bounded LSF.

Lemma 4.5. *A positive operator $T : M(N) \to M(N)$ is rank non-decreasing iff $Cap(T) > 0$.*

Proof. Let us fix an orthonormal basis (unitary matrix) $U = \{u_1, \ldots, u_N\}$ in C^N and associate with a positive operator T the following positive operator:

$$T_U(X) =: \sum_{1 \leqslant i \leqslant N} T(u_i u_i^\dagger) tr(X u_i u_i^\dagger). \tag{30}$$

(In physics terms, T_U represents decoherence with respect to the basis U followed by the application of the map T, i.e. the matrix obtained by applying T_U to matrix X is the same as obtained by applying T to the diagonal restriction of X.)

It is easy to see that a positive operator T is rank non-decreasing iff the operators T_U are rank non-decreasing for all unitary U. (This is because every positive matrix P has a diagonal basis, and T's action on P is the same as that of T_U where U is given by P's diagonal basis. So if there's a P whose rank is decreased by T, then there is a U such that T_U decreases P's rank. Also if there's a T_U that is rank-non-decreasing, it must decrease the rank of some P. Since T_U is decoherence in the basis U followed by T,

and decoherence in a basis is doubly stochastic, hence (cf. Proposition 2.4) rank non-decreasing, T must decrease the rank of the matrix resulting from decohering P in the basis U.)

For fixed U all properties of T_U are defined by the following N-tuple of $N \times N$ positive semidefinite matrices:

$$\mathbf{A}_{T,U} =: (T(u_1 u_1^\dagger), \ldots, T(u_N u_N^\dagger)). \tag{31}$$

Importantly for us, T_U is rank non-decreasing iff the mixed discriminant $M(T(u_1 u_1^\dagger), \ldots, T(u_N u_N^\dagger))$ is strictly positive. Indeed, the operator T_U is rank non-decreasing if and only if the following inequalities hold:

$$Rank\left(\sum_{i\in S} A_i\right) \geqslant |S| \quad \text{for all } S \subset \{1, 2, \ldots, N\}.$$

But these inequalities are equivalent to the inequality $M(T(u_1 u_1^\dagger), \ldots, T(u_N u_N^\dagger)) > 0$ [21,22]. Notice that the operator T is rank non-decreasing if and only if operators T_U are rank non-decreasing for all unitary U.

Define the capacity of $\mathbf{A}_{T,U}$,

$$Cap(\mathbf{A}_{T,U}) =: \inf\left\{Det\left(\sum\nolimits_{1\leqslant i\leqslant N} T(u_i u_i^\dagger)\gamma_i\right) : \gamma_i > 0, \prod\nolimits_{1\leqslant i\leqslant N} \gamma_i = 1\right\}.$$

It is clear from the definitions that $Cap(T)$ is equal to the infimum of $Cap(\mathbf{A}_{T,U})$ over all unitary U.

One of the main results of [21] states that

$$\begin{aligned} M(\mathbf{A}_{T,U}) =: M(T(u_1 u_1^\dagger), \ldots, T(u_N u_N^\dagger)) &\leqslant Cap(\mathbf{A}_{T,U}) \\ &\leqslant \frac{N^N}{N!} M(T(u_1 u_1^\dagger), \ldots, T(u_N u_N^\dagger)). \end{aligned} \tag{32}$$

As the mixed discriminant is a continuous (analytic) functional and the group $SU(N)$ of unitary matrices is compact, we get the next inequality:

$$\min_{U\in SU(N)} M(\mathbf{A}_{T,U}) \leqslant Cap(T) \leqslant \frac{N^N}{N!} \min_{U\in SU(N)} M(\mathbf{A}_{T,U}) \tag{33}$$

But the operator T is rank non-decreasing if and only if operators T_U are rank non-decreasing for all unitary U. Or equivalently, if and only if $\min_{U\in SU(N)} M(\mathbf{A}_{T,U}) > 0$.

Therefore inequality (33) proves that $Cap(T) > 0$ iff positive operator T is rank non-decreasing. □

So, the capacity is a bounded LSF responsible for matching, which proves the next theorem:

Theorem 4.6. 1. *Let $T_n, T_0 = T$ be a trajectory of (OSI), where T is a positive linear operator. Then $DS(T_n)$ converges to zero iff T is rank non-decreasing.*

2. *A positive linear operator $T : M(N) \to M(N)$ is rank non-decreasing iff for all $\varepsilon > 0$ there exists an ε-doubly stochastic operator scaling of T.*

3. *A positive linear operator T is rank non-decreasing iff there exists $\frac{1}{N}$-doubly stochastic operator scaling of T.*

The next theorem generalizes second part of Proposition 4.1 and is proved on almost the same lines as Lemmas 3.2–3.8 in [22].

Theorem 4.7. 1. *There exist non-singular matrices C_1, C_2 such that $S_{C_1,C_2}(T)$ is doubly stochastic iff the infimum in Definition 4.4 is strictly positive and attained. Moreover, if $Cap(T) = Det(T(C))$ where $C \succ 0$, $Det(C) = 1$ then $S_{T(C)^{\frac{-1}{2}},C^{\frac{1}{2}}}(T)$ is doubly stochastic.*

2. *The positive operator T is indecomposable iff the infimum in Definition 4.4 is attained and unique.*

Proof. 1. Suppose that there exist non-singular matrices C_1, C_2 such that $S_{C_1,C_2}(T) =: Z$ is doubly stochastic. Thus $T(X) = C_1^{-1} Z(C_2^{-\dagger} X C_2^{-1}) C_1^{-\dagger}$. It is well known and easy to prove that any doubly stochastic operator Z satisfies the inequality $Det(Z(X)) \geqslant Det(X)$, $X \succ 0$. It follows that

$$Det(T(X)) \geqslant |Det(C_1)|^{-2} |Det(C_2)|^{-2}\, Det(X),\ X \succ 0$$

and

$$Det(T(C_2^{\dagger} C_2)) = |Det(C_1)|^{-2} Det(I) = |Det(C_1)|^{-2} |Det(C_2)|^{-2}\, Det(T(C_2^{\dagger} C_2)).$$

Thus the infimum in Definition 4.4 is positive and attained at $X = |Det(C_2)|^{\frac{2}{N}} C_2^{\dagger} C_2$. Suppose there exists $C \succ 0$, $Det(C) = 1$ such that $T(C) = \inf\{Det(T(X)) : X \succ 0, Det(X) = 1\}$ and $T(C) \succ 0$. We will adapt below the proof of Lemma 3.7 in [22] to this operator setting, using Lagrange multipliers. Consider the following conditional minimization problem:

$$f(X) =: \log(Det(T(X))) \to \min : X \succ 0,\ \log(Det(X)) = 0.$$

Clearly, the positive definite matrix C is a minimum for this problem. Thus, using Lagrange multipliers, we get that when evaluated at C the gradient, $(\nabla f)_C = const\, (\nabla \log(det(\cdot))))_C$. Recall that the gradient $(\nabla \log det(\cdot))_B$ at a non-singular hermitian matrix B is just its inverse B^{-1}. Using the Chain Rule, we get that

$$(\nabla \log(det(T(\cdot))))_C = T^*(T(C)^{-1}),$$

where T^* is the dual to T with respect to the inner product $< X, Y >= tr(XY^{\dagger})$.
Thus, we get that $T^*(T(C)^{-1}) = const\ C^{-1}$. Define now a new scaled operator $Z = S_{T(C)^{\frac{-1}{2}},C^{\frac{1}{2}}}(T)$. Then $Z(I) = I$ and $Z^*(I) = const\, I$. As $tr(Z(I)) = tr(Z^*(I))$ we get that this Z is doubly stochastic.

2. Let us first recall Definition 1.7 from [22]: *An N-tuple $\mathbf{A} = (A_1, \ldots A_N)$ of positive semidefinite $N \times N$ matrices is fully indecomposable if for all $S \subset \{1, \ldots, N\}$, $0 < |S| < N$, $Rank(\sum_{i \in S} A_i) > |S|$.* Consider $N(N-1)$ auxiliary N-tuples $\mathbf{A}^{ij}$, where $i \neq j$ and $\mathbf{A}^{ij}$ is obtained from $\mathbf{A}$ by substituting A_i instead of A_j. Let M_{ij} be the mixed discriminant of the tuple $\mathbf{A}^{ij}$ and $K(\mathbf{A}) = \min_{i \neq j} M_{ij}$. It had been proved in [22] that an N-tuple $\mathbf{A} = (A_1, \ldots A_N)$ of positive semidefinite $N \times N$ matrices is fully

indecomposable iff $K(\mathbf{A}) > 0$. (The quantity $K(\mathbf{A})$ can be viewed as a measure of indecomposability.) Moreover, Lemma 3.7 in [22] proves that if

$$det\left(\sum_{1\leqslant i\leqslant N} x_i A_i\right) \leqslant d;\ d > 0 \text{ and } x_i > 0,\ \prod_{1\leqslant i\leqslant N} x_i = 1,$$

then $\frac{x_i}{x_j} \leqslant \frac{2d}{K(\mathbf{A})}$.

It is obvious that a positive operator T is indecomposable iff for all unitary U the tuples $\mathbf{A}_{T,U}$, defined by (31), are fully indecomposable according to Definition 1.7 in [22]. Clearly, the functional $K(\mathbf{A}_{T,U})$ is continuous on the compact set $SU(N)$ of complex unitary $N \times N$ matrices for it is a minimum of a finite number of continuous functionals. Therefore, we get that a positive operator T is indecomposable iff $K(T) =: \min_{U\in SU(N)} K(\mathbf{A}_{T,U}) > 0$.

Suppose that $X \succ 0,\ Det(X) = 1$ and $Det(T(X)) \leqslant Det(T(I))$; $(\gamma_1 \geqslant \gamma_2 \geqslant \cdots \geqslant \gamma_N)$ are eigenvalues of X and that this X is diagonalized by some unitary matrix $U \in SU(N)$. Using Lemma 3.7 in [22] we get that

$$\frac{\gamma_1}{\gamma_N} \leqslant \frac{2Det(T(I))}{K(\mathbf{A}_{T,U})} \leqslant \frac{2Det(T(I))}{K(T)}.$$

Therefore the infimum in Definition 4.4 can be considered on some compact subset of $\{X \succ 0 : Det(X) = 1\}$ and thus is attained.

The uniqueness part follows fairly directly from Lemma 3.2 in [22]: using the existence part the general indecomposable case can be reduced to the case of indecomposable doubly stochastic operators. It follows from the strict convexity statement in Lemma 3.2 in [22] that if T is doubly stochastic indecomposable operator then $\det(T(X)) > 1$ provided $X \succ 0,\ det(X) = 1$ and $X \neq I$. □

Remark 4.8. Consider an $N \times N$ matrix A with non-negative entries. Similarly to Definition 4.4, define its capacity as follows:

$$Cap(A) = \inf\left\{\prod_{1\leqslant i\leqslant N} (Ax)_i : x_i > 0, 1 \leqslant i \leqslant N;\ \prod_{1\leqslant i\leqslant N} x_i = 1\right\}.$$

Recall that the Kullback–Leibler (KL) divergence between two matrices is defined as

$$KLD(A||B) = \sum_{1\leqslant i,j\leqslant N} B(i,j)\log\left(\frac{B(i,j)}{A(i,j)}\right).$$

It is easy to prove (see, for instance, [23]) that

$$-\log(Cap(A)) = \inf\{KLD(A||B) : B \in D_N\},$$

where D_N is the convex compact set of $N \times N$ doubly stochastic matrices.

Of course, there is a quantum analog of the KL divergence, the so-called quantum Kullback–Leibler divergence. It is not clear whether there exists a similar "quantum" characterization of the capacity of completely positive operators.

It is important not to confuse our capacity with Holevo's capacity, a fundamental quantity in quantum information theory, defined as the capacity of a quantum channel to carry quantum information, or the closely related "Holevo bound" quantity that can be defined for positive operators (or for distributions over density matrices). In fact, our capacity is multiplicative, i.e.

$$\log(Cap(T_1 \otimes T_2)) = \log(Cap(T_1)) + \log(Cap(T_2)).$$

Inequality (20) can be strengthened to the following one:

$$QP(CH(T)) \leqslant N! Cap(T),$$

and $N!$ is an exact constant in this inequality.

If T is separable then $\alpha_N Cap(T) \leqslant QP(CH(T)) \leqslant Cap(T)$, where the positive constant α_N comes from a "third generation" of the Van der Waerden Conjecture [17]. We conjecture that $\alpha_N = \frac{N!}{N^N}$.

5. Polynomial time deterministic algorithm for the Edmonds Problem

Recall that Edmonds' problem is to decide, given a linear subspace $V \subset M(N)$ presented by a k-tuple of rational matrices $\mathbf{A} = (A_1, \ldots, A_k)$ whose span is V, whether V contains a non-singular matrix.

Let us consider the following three properties of BUDM $\rho_{\mathbf{A}}$ associated (as in Eq. (5)) with the k-tuple $\mathbf{A} = (A_1, \ldots, A_k)$. (We will view this $\rho_{\mathbf{A}}$ as Choi's representation of a completely positive operator T, i.e. $\rho_{\mathbf{A}} = CH(T)$.)

P1. $Im(\rho_{\mathbf{A}}) = V = L(A_1, \ldots, A_k)$ contains a non-singular matrix.

P2. The Quantum permanent $QP(\rho_{\mathbf{A}}) > 0$.

P3. Operator T is rank non-decreasing.

Part 2 of Theorem 3.8 proves that P1 $\Leftrightarrow$ P2 and Example 2.8 illustrated that the implication P2 $\Rightarrow$ P3 is strict. It is not clear whether either P1 or P3 can be checked in deterministic polynomial time.

Next, we will describe and analyze a polynomial time deterministic algorithm to check whether P3 holds provided that it is promised that $Im(\rho_{\mathbf{A}})$, viewed as a linear subspace of $M(N)$, has the Edmonds–Rado Property. Or, in other words, that it is promised that P1 $\Leftrightarrow$ P3.

Algorithm 5.1. *The input is a completely positive operator $T : M(N) \to M(N)$ presented by BUDM $\rho_{\mathbf{A}}$ with integer entries. The maximum magnitude of the entries is* $\mathbf{M}$. *The output is "NO" if $QP(\rho_{\mathbf{A}}) = 0$ and "YES" otherwise.*

Step 1. If either $T(I)$ or $T^*(I)$ is singular then output "NO." (Notice that if $T(I) \succ 0$ and $T^*(I) \succ 0$ then all along the trajectory of (OSI) ($T_0 = T, n \geqslant 0$), also $T_n(I) \succ 0$ and $T_n^*(I) \succ 0$.)

Step 2. Compute $L \approx 3N(N \ln(N) + N(\ln(N) + \ln(\mathbf{M}))$.

If $DS(T_L) \leqslant \frac{1}{N}$ then output "YES".

If $DS(T_L) > \frac{1}{N}$ then output "NO".

The "NO" in Step 1 is obvious. The "YES" in Step 2 follows from Theorem 4.6. The only thing which needs further justification is the "NO" in Step 2.

Let $L =: \min\{n : DS(T_n) \leqslant \frac{1}{N}\}$ where the initial T_0 presented by BUDM $\rho_{\mathbf{A}}$ with $QP(\rho_{\mathbf{A}}) > 0$. Then necessarily $QP(\rho_{A,B}) \geqslant 1$ for it is an integer number. Also, $Det(T_0(I)) \leqslant (\mathbf{MN})^{\mathbf{N}}$ by the Hadamard's

inequality. Thus

$$QP(CH(T_1)) = \frac{QP(CH(T))}{Det(T_0(I))} \geqslant (\mathbf{MN})^{-\mathbf{N}}.$$

Each nth iteration ($n \leqslant L$) after the first one will multiply the Quantum permanent by $Det(X)^{-1}$, where $X \succ 0$, $tr(X) = N$ and $tr((X-I)^2) > \frac{1}{N}$. Using Lemma 3.3 from [25], $Det(X)^{-1} \geqslant (1 - \frac{1}{3N})^{-1} =: \delta$. Putting all this together, we get the following upper bound on L, the number of steps in (OSI) to reach the "boundary" $DS(T_n) \leqslant \frac{1}{N}$:

$$\delta^L \leqslant \frac{QP(CH(T_L))}{(\mathbf{MN})^{-\mathbf{N}}}. \tag{34}$$

It follows from (20) that $QP(CH(T_L)) \leqslant N!$

Taking logarithms we get that

$$L \leqslant \approx 3N(N\ln(N) + N(\ln(N) + \ln(\mathbf{M})). \tag{35}$$

Thus L is polynomial in the dimension N and the number of bits $\log_2(\mathbf{M})$.

To finish our analysis, we need to evaluate the complexity of each step of (OSI). Recall that $T_n(X) = L_n(T(R_n^\dagger X R_n))L_n^\dagger$ for some non-singular matrices L_n and R_n, that $T_n(I) = L_n(T(R_n^\dagger R_n))L_n^\dagger$ and that $T_n^*(I) = R_n(T^*(L_n^\dagger L_n))R_n^\dagger$. To evaluate $DS(T_n)$ we need to compute $tr((T_n^*(I) - I)^2)$ for odd n and $tr((T_n(I) - I)^2)$ for even n. Define $P_n = L_n^\dagger L_n$, $Q_n = R_n^\dagger R_n$. Clearly, the matrix $T_n(I)$ is similar to $P_n T(Q_n)$, and $T_n^*(I)$ is similar to $Q_n T^*(P_n)$. As traces of similar matrices are equal, to evaluate $DS(T_n)$ it is sufficient to compute the matrices P_n, Q_n.

It follows from the definition of (OSI) that $P_{n+1} = P_n = (T(Q_n))^{-1}$ for odd n and $Q_{n+1} = Q_n = (T^*(P_n))^{-1}$ for even $n \geqslant 2$. This gives the following recursive algorithm, which we will call Rational Operator Sinkhorn's iterative scaling (ROSI):
$P_{2(k+1)+1} = (T((T^*(P_{2(k)+1}))^{-1}))^{-1}$, $P_1 = (T(I))^{-1}$;
$Q_{2(k+1)} = (T^*((T(P_{2(k)}))^{-1}))^{-1}$, $Q_0 = I$.

Notice that the original definition of (OSI) requires computation of an operator square root. It can be replaced by the Cholesky factorization, which still requires computing scalar square roots. But our final algorithm is rational with $O(N^3)$ arithmetic operations per iteration!

Remark 5.1. Algorithms of this kind for the "classical" matching problem appeared independently in [25,23]. In the "classical" case they are just another, conceptually simple, but far from optimal, poly-time algorithm to check whether a perfect matching exists. In this general Edmonds Problem setting, is our Operator Sinkhorn's iterative scaling based approach the only possibility?

5.1. Controlling the bit-size

The bit-size of the entries of the matrices P_i and Q_i might grow exponentially. On the other hand, our algorithm still produces a correct answer if :

1. We round the entries of the matrices P_i and Q_i in such a way that the estimate of $DS(T_L)$ is accurate up to absolute error $\frac{1}{2N}$.

   Rounding here amounts in introducing errors : $\overbrace{P_i} = P_i + \delta_i$, $\overbrace{Q_i} = P_i + \varepsilon_i$

2. We check the condition "$DS(T_L) \leqslant \frac{1}{2N}$" (instead of the condition "$DS(T_L) \leqslant \frac{1}{N}$")
3. We run the algorithm a bit, i.e. twice, longer.

In order to make this idea work we need that $|\delta_i|, |\varepsilon_i| \geqslant 2^{-poly(N, \ln(\mathbf{M}))}$. The problem we face here is essentially a sensitivity analysis of the following dynamical system:

$$X_{n+1} = (T((T^*(X_n))^{-1}))^{-1}, X_0 = (T(I))^{-1}.$$

The next theorem, proved in Appendix D, shows that such a polynomial $poly(N, \ln(\mathbf{M}))$ does exist.

Theorem 5.2. *Consider* (*not rounded*) *Rational Operator Sinkhorn's iterative scaling* (*ROSI*):

$$P_{2(k+1)+1} = (T((T^*(P_{2(k)+1}))^{-1}))^{-1}, P_1 = (T(I))^{-1},$$
$$Q_{2(k+1)} = (T^*((T(P_{2(k)}))^{-1}))^{-1}, Q_0 = I.$$

Consider also the rounded Rational Operator Sinkhorn's iterative scaling (RROSI), i.e. the recursion

$$\overbrace{P_{2(k+1)+1}} = (T((T^*(\overbrace{P_{2(k)+1}}))^{-1}))^{-1} + \delta_k, \overbrace{P_1} = (T(I))^{-1},$$
$$\overbrace{Q_{2(k+1)}} = (T^*((T(\overbrace{P_{2(k)}}))^{-1}))^{-1} + \varepsilon_k, \overbrace{Q_0} = I,$$

There exists a polynomial $poly(N, \ln(\mathbf{M}))$ *such that if the norms (i.e. the largest magnitudes of the eigenvalues) of hermitian matrices* δ_k, ε_k *satisfy the inequalities*

$$||\delta_k|| \leqslant 2^{-poly(N, \ln(\mathbf{M}))}, ||\varepsilon_k|| \leqslant 2^{-poly(N, \ln(\mathbf{M}))},$$

then $|\overbrace{DS(T_n)} - DS(T_n)| \leqslant \frac{1}{2N}$ *for all* $0 \leqslant n \leqslant N^2(N \ln(N) + N(\ln(N) + \ln(\mathbf{M}))$. *(Here* $\overbrace{DS(T_n)} = tr((\overbrace{P_n}$ $T(\overbrace{Q_n}) - I))^2) + tr((\overbrace{Q_n}\, T^*(\overbrace{P_n}) - I))^2)$ *and* $DS(T_n) = tr((P_n T(Q_n) - I)^2) + tr((Q_n T(P_n) - I)^2)$.*)*

Remark 5.3. There is nothing special about the quantity $N^2(N \ln(N) + N(\ln(N) + \ln(\mathbf{M}))$, it just large enough, i.e. larger than twice the number of iterations of (not rounded) Algorithm 5.1.

6. Weak Membership Problem for the convex compact set of normalized bipartite separable density matrices is NP-HARD

One of the main research activities in Quantum Information Theory is a search for an "operational" criterion for separability. We will show in this section that, in a sense defined below, the problem is NP-HARD even for bipartite normalized density matrices provided that each part is large (each "particle" has large number of levels). First, we need to recall some basic notions from computational convex geometry.

6.1. Algorithmic aspects of convex sets

We will follow [14].

Definition 6.1. A proper (i.e. with non-empty interior) convex set $K \subset R^n$ is called well-bounded a-centered if there exist a rational vector $a \in K$ and positive (rational) numbers r, R such that $B(a, r) \subset K$

and $K \subset B(a, R)$ (here $B(a, r) = \{x : \|x - a\| \leqslant r\}$ and $\|.\|$ is a standard euclidean norm in R^n). The encoding length of such a convex set K is

$$< K >= n + < r > + < R > + < a >,$$

where $< r >, < R >, < a >$ are the number of bits of corresponding rational numbers and rational vector. Following [14] we define $S(K, \delta)$ as a union of all δ-balls with centers belonging to K ; and $S(K, -\delta) = \{x \in K : B(x, \delta) \subset K\}$.

Definition 6.2. The Weak Membership Problem
$(WMEM(K, y, \delta))$ is defined as follows:
Given a rational vector $y \in R^n$ and a rational number $\delta > 0$ either
(i) assert that $y \in S(K, \delta)$, or
(ii) assert that $y \notin S(K, -\delta)$.

The Weak Validity Problem $(WVAL(K, c, \gamma, \delta))$ is defined as follows:
Given a rational vector $c \in R^n$, rational number γ and a rational number $\delta > 0$ either
(i) assert that $< c, x >=: c^T x \leqslant \gamma + \delta$ for all $x \in S(K, -\delta)$, or
(ii) assert that $c^T x \geqslant \gamma - \delta$ for some $x \in S(K, \delta)$.

Remark 6.3. Define $M(K, c) =: \max_{x \in K} < c, x >$. It is easy to see that

$$\begin{aligned} M(K, c) &\geqslant M(S(K, -\delta), c) \geqslant M(K, c) - \|c\|\delta\frac{R}{r}, \\ M(K, c) &\leqslant M(S(K, \delta), c) \leqslant M(K, c) + \|c\|\delta \end{aligned}$$

Recall that the seminal Yudin–Nemirovski theorem [32,14] implies that if there exists a deterministic algorithm solving $WMEM(K, y, \delta)$ in $Poly(< K > + < y > + < \delta >)$ steps then there exists a deterministic algorithm solving $WVAL(K, c, \gamma, \delta)$ in $Poly(< K > + < c > + < \delta > + < \gamma >)$ steps.

Let us denote as $SEP(M, N)$ the compact convex set of separable density matrices $\rho_{A,B} : C^M \otimes C^N \to C^M \otimes C^N$, $tr(\rho_{A,B}) = 1$, $M \geqslant N$. Recall that

$$\begin{aligned} &SEP(M, N) \\ &\quad = CO(\{xx^\dagger \otimes yy^\dagger : x \in C^M, y \in C^N; \|x\| = \|y\| = 1\}), \end{aligned}$$

where $CO(X)$ stands for the convex hull generated by a set X.

Our goal is to prove that the Weak Membership Problem for $SEP(M, N)$ is NP-HARD. As we are going to use the Yudin–Nemirovski theorem, it is sufficient to prove that $WVAL(SEP(M, N), c, \gamma, \delta)$ is NP-HARD with respect to the complexity measure $(M + < c > + < \delta > + < \gamma >)$ and to show that $< SEP(M, N) >$ is polynomial in M.

6.2. Geometry of $SEP(M, N)$

First, $SEP(M, N)$ can be viewed as a compact convex subset of the hyperplane in R^{D^2}, $D =: NM$. The standard euclidean norm in $R^{N^2M^2}$ corresponds to the Frobenius norm for density matrices, i.e. $\|\rho\|_F = tr(\rho\rho^\dagger)$. The matrix $\frac{1}{NM}I \in SEP(M, N)$ and $\|\frac{1}{NM}I - xx^\dagger \otimes yy^\dagger\|_F = \sqrt{\frac{D-1}{D}} < 1$ for all norm one vectors x, y. Thus $SEP(M, N)$ is covered by the ball $B(\frac{1}{NM}I, \sqrt{\frac{D-1}{D}})$.

The following result was recently proved in [20].

Theorem 6.4. *Let Δ be a block hermitian matrix as in* (5). *If $tr(\Delta) = 0$ and $\|\Delta\|_F \leqslant \sqrt{\frac{1}{D(D-1)}}$ then the block matrix $\frac{1}{D}I + \Delta$ is separable.*

Summarizing, we get that for $D = MN$

$$B\left(\frac{1}{D}I, \sqrt{\frac{1}{D(D-1)}}\right) \subset SEP(M, N) \subset B\left(\frac{1}{D}I, \sqrt{\frac{D-1}{D}}\right)$$

(balls are restricted to the corresponding hyperplane) and conclude that $< SEP(M, N) > \leqslant Poly(MN)$. It is left to prove that $WVAL(SEP(M, N), c, \gamma, \delta)$ is NP-HARD with respect to the complexity measure $(MN + < c > + < \delta > + < \gamma >)$.

6.3. Proof of hardness

Let us consider the following hermitian block matrix:

$$C = \begin{pmatrix} 0 & A_1 & \dots & A_{M-1} \\ A_1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ A_{M-1} & 0 & \dots & 0 \end{pmatrix}, \tag{36}$$

i.e. its (i, j) blocks are zero if either $i \neq 1$ or $j \neq 1$ and $(1, 1)$ block is also zero; $A_1, \dots, A_{M-1}$ are real symmetric $N \times N$ matrices.

Proposition 6.5.

$$\max_{\rho \in SEP(M,N)} (tr(C\rho))^2 = \max_{y \in R^N, \|y\|=1} \sum\nolimits_{1 \leqslant i \leqslant M-1} (y^T A_i y)^2.$$

Proof. First, by linearity and the fact that the set of extreme points $Ext(SEP(M, N))$ is equal to

$$\{xx^\dagger \otimes yy^\dagger : x \in C^M, y \in C^N; \|x\| = \|y\| = 1\},$$

we get that

$$\max_{\rho\in SEP(M,N)} tr(C\rho)$$
$$= \max_{xx^\dagger\otimes yy^\dagger: x\in C^M, y\in C^N; \|x\|=\|y\|=1} tr(C(xx^\dagger\otimes yy^\dagger)).$$

But $tr(C(yy^\dagger\otimes xx^\dagger)) = tr(A(y)xx^\dagger)$, where the real symmetric $M\times M$ matrix $A(y)$ is defined as follows:

$$A(y) = \begin{pmatrix} 0 & a_1 & \dots & a_{M-1} \\ a_1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ a_{M-1} & 0 & \dots & 0 \end{pmatrix},$$

$a_i = tr(A_i yy^\dagger)$, $1\leqslant i\leqslant M-1$.
Thus

$$\max_{\rho\in SEP(M,N)} tr(C\rho)$$
$$= \max_{yy^\dagger\otimes xx^\dagger: x\in C^M, y\in C^N; \|x\|=\|y\|=1} tr(C(xx^\dagger\otimes yy^\dagger))$$
$$= \max_{\|y\|=1} \lambda_{max} A(y).$$

(Above, $\lambda_{max}A(y)$ is the maximum eigenvalue of $A(y)$.)

It is easy to see $A(y)$ has only two real non-zero eigenvalues $(d,-d)$, where $d^2 = \sum_{1\leqslant i\leqslant M-1}(tr(A_i yy^\dagger))^2$. As A_i, $1\leqslant i\leqslant N-1$ are real symmetric matrices we finally get that

$$\max_{\rho\in SEP(M,N)} (tr(C\rho))^2 = \max_{y\in R^N, \|y\|=1} \sum_{1\leqslant i\leqslant N-1} (y^T A_i y)^2. \qquad \square$$

Proposition 6.5 and Remark 6.3 demonstrate that in order to prove NP-HARDness of $WVAL(SEP(M,N), c, \gamma, \delta)$ with respect to the complexity measure $M+<c>+<\delta>+<\gamma>$ it is sufficient to prove that the following problem is NP-HARD:

Definition 6.6. (RSDF problem). Given k $l\times l$ real rational symmetric matrices $(A_i, 1\leqslant i\leqslant l)$ and rational numbers (γ,δ) to check whether

$$\gamma+\delta \geqslant \max_{x\in R^l, \|x\|=1} f(x) \geqslant \gamma-\delta,\ f(x) =: \sum_{1\leqslant i\leqslant l} (x^T A_i x)^2$$

with respect to the complexity measure

$$\left(lk + \sum_{1\leqslant i\leqslant l} <A_i> + <\delta> + <\gamma> \right).$$

It was shown in [6], by a reduction from KNAPSACK, that the RSDF problem is NP-HARD provided $k\geqslant \frac{l(l-1)}{2}+1$.

We summarize all this in the following theorem:

Theorem 6.7. *The Weak Membership Problem for $SEP(M, N)$ is NP-HARD if $N \leqslant M \leqslant \frac{N(N-1)}{2} + 2$.*

Remark 6.8. It is an easy exercise to prove that BUDM $\rho_{A,B}$ written in block form (3) is real separable iff it is separable and all the blocks in (3) are real symmetric matrices. It follows that, with obvious modifications, Theorem 6.7 is valid for real separability too.

Construction (37) was inspired by Arkadi Nemirovski's proof of the NP-HARDness of checking the positivity of a given operator [27].

7. Concluding remarks

Many ideas of this paper were initiated in [21]. The main technical result in a recent breakthrough in Communication Complexity [13] is a rediscovery of particular, rank one, case of a general, matrix tuples scaling, result proved in [21] with a much simpler proof than in [13]. Perhaps this paper will produce something new in Quantum Communication Complexity.

We still do not know whether there is a deterministic poly-time algorithm to check if a given completely positive operator is rank non-decreasing. This question is related to lower bounds on $Cap(T)$ provided that Choi's representation $CH(T)$ is an integer semidefinite matrix. Another interesting open question is about the "power" of ERP for Edmonds' problem over finite fields.

Theorem 6.7 together with other results from our paper gives a new, classical complexity based, insight into the nature of quantum entanglement and, in a sense, closes a long line of research in Quantum Information Theory. Also, this paper suggests a new way to look at "the worst entangled" bipartite density matrices (or completely positive operators). For instance, the operator Sk_3 from Example 2.8 seems to be "the worst entangled" and it is not surprising that it appears in many counterexamples.

The G-norm defined in (2) appears in this paper mainly because of formula (24). It is called by some authors [33] Bombieri's norm (see also [5,29,4]). Also, the G-norm arises naturally in quantum optics and the study of quantum harmonic oscillators. This norm satisfies some remarkable properties [5,29] which, we think, can be used in quantum/linear optics computing research.

Combining formulas (23) and (24), one gets an unbiased non-negative valued random estimator for quantum permanents of bipartite unnormalized density matrices. A particular case of this construction is a simple unbiased non-negative valued random estimator for permanents of positive semidefinite matrices (Corollary A.1 and formula (43)). But, as indicated in [15], it behaves rather badly for the entangled bipartite unnormalized density matrices. On the other hand, there is hope, depending on a proof of a "third generation" of van der Waerden conjecture [12,11,21,16], to have even a deterministic polynomial time algorithm to approximate within a simply exponential factor quantum permanents of separable unnormalized bipartite density matrices (more details on this matter can be found in [17]).

Acknowledgements

It is my great pleasure to thank my LANL colleagues Manny Knill and Howard Barnum. Thanks to Marek Karpinski and Alex Samorodnitsky for their comments on the earlier version of this paper. I thank Arkadi Nemirovski for many enlightening discussions.

Finally, I am greatly indebted to the anonymous reviewer for amazingly careful reading of this paper.

Appendix A. Proof of Theorem 3.8 and a permanental corollary

The main goal of this appendix is a "direct proof" of formula (24). A much shorter probabilistic proof is presented in Appendix **C** .

Proof of formula (23). It is sufficient to prove that for any monomial

$$\frac{1}{\pi^k}\int\ldots\int |z_1^{r_1}\ldots z_k^{r_k}|^2 e^{-(x_1^2+y_1^2)}\ldots e^{-(x_k^2+y_k^2)}dx_1\,dy_1\ldots dx_k\,dy_k = r_1!r_2!\ldots r_k!$$
$$\times(z_l = x_l + iy_l,\ 1\leqslant l\leqslant k). \tag{A.1}$$

And that distinct monomials are orthogonal, i.e.

$$\int\ldots\int (z_1^{r_1}\ldots z_k^{r_k}\overline{z_1^{h_1}\ldots z_k^{h_k}})e^{-(x_1^2+y_1^2)}\ldots e^{-(x_k^2+y_k^2)}dx_1\,dy_1\ldots dx_k\,dy_k = 0 (r\neq h). \tag{A.2}$$

Notice that both $2k$-dimensional integrals (37) and (38) are products of corresponding 2-dimensional integrals. Thus (37) is reduced to the fact that

$$\frac{1}{\pi}\int\int (x_1^2+y_1^2)^{2r_1}e^{-(x_1^2+y_1^2)}\,dx_1\,dy_1 = r_1!.$$

Using polar coordinates in a standard way, we get that

$$\frac{1}{\pi}\int\int (x_1^2+y_1^2)^{2r_1}e^{-(x_1^2+y_1^2)}\,dx_1\,dy_1 = \int_0^\infty R^{r_1}e^{-R}dR = r_1!.$$

Similarly (38) is reduced to

$$\int\int (x_1+iy_1)^m(x_1^2+y_1^2)^k e^{-(x_1^2+y_1^2)}\,dx_1\,dy_1 = 0,$$

where m is positive integer and k is non-negative integer.
But

$$\int\int (x_1+iy_1)^m(x_1^2+y_1^2)^k e^{-x_1^2+y_1^2}dx_1dy_1 = \int_0^\infty R^{2k+m}e^{-R^2}\left(\int_0^{2\pi} e^{-im\phi}d\phi\right)dR = 0. \qquad \square$$

Proof of formula (24). First, let us recall how coefficients of $det(\sum_{1\leqslant i\leqslant k} x_i A_i)$ can be expressed in terms of the corresponding mixed discriminants. Let us associate a vector $r\in I_{k,N}$ an N-tuple of $N\times N$ complex matrices $\mathbf{B}_r$ consisting of r_i copies of $A_i (1\leqslant i\leqslant k)$.
Notice that

$$\mathbf{B}_r = (B_1,\ldots,B_N);\ B_i\in\{A_1,\ldots,A_k\},\quad 1\leqslant i\leqslant k.$$

It is well known and easy to check that for this particular determinantal polynomial its coefficients satisfy the following identities:

$$a_{r_1,\ldots,r_k} = \frac{M(\mathbf{B}_r)}{r_1! r_2! \ldots r_k!}, \quad r = (r_1, \ldots, r_k) \in I_{k,N}, \tag{A.3}$$

where $M(\mathbf{B}_r)$ is the mixed discriminant of the tuple $\mathbf{B}_r$.
We already defined mixed discriminants by two equivalent formulas (12), (13). The next equivalent definition is handy for our proof:

$$M(B_1, \ldots B_N) = \sum_{\sigma \in S_N} det([B_1(e_{\sigma(1)}) | B_2(e_{\sigma(2)}) | \ldots | B_N(e_{\sigma(N)})]). \tag{A.4}$$

In formula(40) above, $(e_1, \ldots, e_N)$ is a canonical basis in C^N, and for a $N \times N$ complex matrix B a column vector $B(e_i)$ is an ith column of B.

We will use in this proof three basic elementary facts about mixed discriminants. First is "local additivity", i.e.

$$M(A_1 + B, A_2, \ldots, A_N) = M(A_1, A_2, \ldots, A_N) + M(B, A_2, \ldots, A_N).$$

Second is permutation invariance, i.e.

$$M(A_1, A_2, \ldots, A_N) = M(A_{\tau(1)}, A_{\tau(2)}, \ldots, A_{\tau(N)}), \tau \in S_N.$$

And the third one is easy formula for the rank one case:

$$M(x_1 y_1^T, \ldots, x_N y_N^T) = det(x_1 y_1^T + \cdots + x_N y_N^T),$$

where $(x_i, y_i; 1 \leqslant i \leqslant N)$ are N-dimensional complex column-vectors.

Recall that the blocks of $\rho_{\mathbf{A}}$ are defined as

$$A_{i,j} = \sum_{1 \leqslant l \leqslant k} A_l e_i e_j^\dagger A_l^\dagger, \quad 1 \leqslant i, j \leqslant N.$$

Let us rewrite formula (14) as follows:

$$QP(\rho) =: \frac{1}{N!} \sum_{\sigma, \tau \in S_N} (-1)^{sign(\sigma)} M(A_{\tau(1),\sigma(1)}, \ldots, A_{\tau(N),\sigma(N)}). \tag{A.5}$$

Using this formula (41) we get the following expression for quantum permanent of bipartite density matrix $\rho_{\mathbf{A}}$ using "local" additivity of mixed dicriminant in each matrix component:

$$QP(\rho_{\mathbf{A}}) = \frac{1}{N!} \sum_{t_1,\ldots,t_N} \sum_{\tau_1, \tau_2 \in S_N} M(A_{t_1} e_{\tau_1(1)} e_{\tau_2(1)}^\dagger A_{t_1}^\dagger, \ldots, A_{t_N} e_{\tau_1(N)} e_{\tau_2(N)}^\dagger A_{t_N}^\dagger).$$

Using the rank one case formula (A) above and formula (40), we get that

$$\sum_{\tau_1, \tau_2 \in S_N} M(A_{t_1} e_{\tau_1(1)} e_{\tau_2(1)}^\dagger A_{t_1}^\dagger, \ldots, A_{t_N} e_{\tau_1(N)} e_{\tau_2(N)}^\dagger A_{t_N}^\dagger) = |M(A_{t_1}, \ldots, A_{t_N})|^2.$$

This formula gives the following, intermediate, identity:

$$QP(\rho_{\mathbf{A}}) = \frac{1}{N!} \sum_{t_1,\ldots,t_N} |M(A_{t_1}, \ldots, A_{t_N})|^2. \tag{A.6}$$

What is left is to "collect" in (42), using invariance of mixed discriminants respect to permutations, all occurences of $M(\mathbf{B}_r)$ (as defined in (39)), where $r = (r_1, \ldots, r_k) \in I_{k,N}$.
It is easy to see that this number $N(r_1, \ldots, r_k)$ of occurences of $M(\mathbf{B}_r)$ is equal to the coefficient of monomial $x_1^{r_1} x_2^{r_2} \ldots x_k^{r_k}$ in the polynomial $(x_1 + \cdots + x_k)^N$. In other words, $N(r_1, \ldots, r_k) = \frac{N!}{r_1! \ldots r_k!}$, which finally gives that

$$QP(\rho_{\mathbf{A}}) = \sum_{r \in I_{k,N}} \frac{|M(\mathbf{B}_r)|^2}{r_1! \ldots r_k!}.$$

Using formula (39) for coefficients of determinantal polynomial $det(\sum_{1 \leqslant i \leqslant k} x_i A_i)$ we get that

$$\| P_{\mathbf{A}} \|_G^2 = \sum_{(r_1,\ldots,r_k) \in I_{k,N}} |a_{r_1,\ldots,r_k}|^2 r_1! r_2! \ldots r_k! = QP(\rho_{\mathbf{A}}). \qquad \square$$

Putting Parts 1 and 2 together we get in the next corollary a formula expressing permanents of positive semidefinite matrices as squares of G-norms of multilinear polynomials. A particular, rank two case, of this formula was (implicitly) discovered in [29].

Corollary A.1. *Consider complex positive semidefinite $N \times N$ matrix $Q = DD^{\dagger}$, where a "factor" D is $N \times M$ complex matrix. Define a complex gaussian vector $z = D\xi$, where ξ is an M-dimensional complex gaussian vector as in theorem 3.8.*
The following formula provides unbiased non-negative valued random estimator for $Per(Q)$:

$$Per(Q) = E_{\xi_1,\ldots,\xi_N}(|z_1|^2 \ldots |z_N|^2). \tag{A.7}$$

Proof. Consider the following m-tuple of complex $N \times N$ matrices:

$$\mathbf{Diag} = (Diag_1, \ldots, Diag_m);\ Diag_j = Diag(D(1, j), \ldots, D(N, j)), \quad 1 \leqslant j \leqslant M.$$

Then $P_{\mathbf{Diag}}(x_1, \ldots, x_m) = \prod_{1 \leqslant i \leqslant N} (Dx)_i$, where $(Dx)_i$ is ith component of vector Dx. Thus Part 1 of theorem 3.8 gives that $\| P_{\mathbf{Diag}} \|_G^2 = E_{z_1,\ldots,z_N}(|z_1|^2 \ldots |z_N|^2)$.

It is easy to see that the block representation of the bipartite density matrix $\rho_{\mathbf{Diag}}$ associated with m-tuple **Diag** is as follows:

$$\rho_{\mathbf{Diag}} = \begin{pmatrix} A_{1,1} & A_{1,2} & \ldots & A_{1,N} \\ A_{2,1} & A_{2,2} & \ldots & A_{2,N} \\ \ldots & \ldots & \ldots & \ldots \\ A_{N,1} & A_{N,2} & \ldots & A_{N,N} \end{pmatrix}, A_{i,j} = Q(i, j) e_i e_j^T.$$

Therefore $QP(\rho_{\mathbf{Diag}}) = Per(Q)$. Now Part 2 of theorem 3.8 gives that

$$Per(Q) = QP(\rho_{\mathbf{Diag}}) = \|P_{\mathbf{Diag}}\|_G^2 = E_{z_1,\ldots,z_N}(|z_1|^2 \ldots |z_N|^2). \qquad \square \tag{A.8}$$

Remark A.2. Corollary (A.1) together with a remarkable supermultiplicative inequality for the G-norm [5,29] give a completely new look at many nontrivial permanental inequalities, such Leib's famous inequality [26], etc., and allow new correlational inequalities for analytic functions of complex gaussian vectors and new ("short") characterizations of independence of analytic functions of complex gaussian vectors. More on this will be described in [18].

Appendix B. Wick formula

In the next theorem we recall the famous Wick formula (see, for instance, [34]).

Theorem B.1. *Consider a complex $2N \times M$ matrix A and a real M-dimensional gaussian vector x with zero mean and covariance matrix $E(xx^T) = I$. Define $(y_1, \ldots, y_{2N})^T = Ax$. Then the following Wick formula holds*:

$$W(A) =: E\left(\prod_{1 \leqslant i \leqslant 2N} y_i\right) = Haf(AA^T), \tag{B.1}$$

where the hafnian $Haf(B)$ of $2N \times 2N$ matrix B is defined as follows:

$$Haf(B) = \sum_{1 \leqslant p_1 < p_2 < \ldots p_N;\, p_1 < q_1, \ldots, p_N < q_N \leqslant 2N} \;\prod_{1 \leqslant i \leqslant N} B(p_i, q_i). \tag{B.2}$$

Let us show how formula (43) follows from (45).

Proposition B.2. *Suppose that the complex $N \times M$ matrix D in Theorem 1.4 can be written as $D = C + iB$. Consider the following complex $2N \times 2M$ matrix A*:

$$\sqrt{2}A = \begin{pmatrix} C + iB & iC - B \\ C - iB & -B - iC \end{pmatrix}.$$

Then

$$AA^T = \begin{pmatrix} 0 & DD^\dagger \\ DD^\dagger & 0 \end{pmatrix},$$

and $W(A) = E_{z_1,\ldots,z_N}(|z_1|^2 \ldots |z_N|^2)$, where the expression $E_{z_1,\ldots,z_N}(|z_1|^2 \ldots |z_N|^2)$ is the same as in Corollary A.1.

As it is easy to see that $Haf(AA^T) = Per(DD^\dagger)$, using the Wick formula (45) we reprove formula (43).

Summarizing, we can say at this point that formula (43) is essentially a different way to write Wick formula. (We thank A.Barvinok for pointing out this observation and Ref. [34]).

On the other hand, formula (43) is a direct corollary of formula (24) for the case of tuples of diagonal matrices. It is easy to see that one can also consider upper triangular matrices. More generally consider the following group action on tuples of square complex matrices $\mathbf{A} = (A_1, \ldots, A_k)$:

$$\mathbf{A}_{X,Y} = (XA_1Y, \ldots, XA_kY), det(XY) = det(X)det(Y) = 1. \tag{B.3}$$

As

$$P_{\mathbf{A}_{X,Y}}(x_1, \ldots, x_k) = det\left(\sum_{1 \leqslant i \leqslant k} x_i XA_iY\right) = det(X)det(Y)P_{\mathbf{A}}(x_1, \ldots, x_k)$$

this group action does not change the corresponding determinantal polynomial. Finally, it follows that the Wick formula is a particular case of formula (24) when there exist two matrices X and Y such that $det(XY) = det(X)det(Y) = 1$ and the matrices $(XA_1Y, \ldots, XA_kY)$ are all upper triangular, or, in Lie-algebraic language, when there exist two non-singular matrices X and Y such that the Lie algebra generated by $(XA_1Y, \ldots, XA_kY)$ is solvable. It seems reasonable to predict that formula (24) might be of good use in the combinatorics described in [34].

Appendix C. Short probabilistic proof of formula (24)

C.1. Hilbert space of analytical functions

Consider a Hilbert space $L_{k,G}$ of analytic functions

$$f(x_1, x_2, \ldots, x_k) = \sum_{(r_1,\ldots,r_k)} a_{r_1,\ldots,r_k} x_1^{r_1} x_2^{r_2} \ldots x_k^{r_k},$$

where the G- inner product is defined as

$$< f, g >_G = \sum_{(r_1,\ldots,r_k)} a_{r_1,\ldots,r_k} \overline{b_{r_1,\ldots,r_k}} r_1! \ldots r_k! \tag{C.1}$$

It is easy to see that $L_{k,G}$ is a closed proper subspace of $L_2(C^k, \mu)$, where μ is a gaussian measure on C^k, i.e. its density function

$$p(z) = \frac{1}{\pi^k} e^{-|z|^2}.$$

Notice that $e^{<x,y>}, x, y \in C^k$ is a reproducing kernel for $L_{k,G}$, i.e. define $g_y(x) =: e^{<x,y>}$; then $g_y \in L_{k,G}$ and $< f, g_y >_G = f(y)$.

Let $f : C^k \to C$ be a complex-valued functional in k complex variables and $A \in M(k)$ be complex $k \times k$ matrix. We denote below a functional $f(AZ), Z \in C^N$ as $f \circ A$ (we view Z as a k-dimensional column vector).

Proposition C.1. *Suppose that $f, g \in L_2(C^k, \mu)$ and the matrix $U : C^k \to C^k$ is unitary, i.e. $UU^* = I$. Then*

$$< f(Ux), g >_{L_2(C^k,\mu)} = < f, g(U^*x) >_{L_2(C^k,\mu)} .$$

Proof. This is just a reformulation of a well known obvious fact that $p(z) = p(Uz)$ ($e^{-|z|^2} = e^{-|Uz|^2}$) for unitary U. □

Lemma C.2. *Let $P(x_1, x_2, \ldots, x_k)$ be a homogeneous polynomial of total degree N and $g \in L_2(C^k, \mu)$. Then for any matrix $A : C^k \to C^k$ the following identity holds*:

$$< P \circ A, g >_{L_2(C^k,\mu)} = < P, g \circ A^* >_{L_2(C^k,\mu)} . \tag{C.2}$$

Proof. First, there is an unique decomposition $g = Q + \delta$, where $Q(x_1, x_2, \ldots, x_k)$ is a homogeneous polynomial of total degree N and $< R, \delta >_{L_2(C^k,\mu)} = 0$ for any homogeneous polynomial R of total degree N. As $P(Ax)$ is a homogeneous polynomial of total degree N for all A thus $< P(Ax), \delta >_{L_2(C^k,\mu)} \equiv 0$.

It is left to prove (49) only when g is a homogeneous polynomial of total degree N. We already know that (49) holds for unitary A. Also, because of formula (23), in this homogeneous case (49) holds for diagonal A. To finish the proof, we use the singular value decomposition $A = VDiagU$, where U, V are unitary and *Diag* is a diagonal matrix with non-negative entries. □

Remark C.3. The homogeneous part of Lemma C.2 has been proved in [29] using the fact that the linear space of homogeneous polynomials of total degree N is spanned by N powers of linear forms.

C.2. Unbiased estimators for quantum permanents

Remark C.4. Consider a four-dimensional tensor $\rho(i_1, i_2, i_3, i_4)$, $1 \leqslant i_1, i_2, i_3, i_4 \leqslant N$. One can view it as a block matrix as in (3), where the blocks are defined by

$$A_{i_1,j_1} =: \{\rho(i_1, i_2, j_1, j_2) : 1 \leqslant i_2, j_2 \leqslant N\}, \quad 1 \leqslant i_1, j_1 \leqslant N$$

We also can permute indices: $\rho(i_{\pi(1)}, i_{\pi(2)}, i_{\pi(3)}, i_{\pi(4)})$, and get another block matrix. The main point is that it follows from formula (17) that a permutation of indices does not change the quantum permanent $QP(\rho)$. In what follows below, we will use the following simple and natural trick: permute indices and use mixed discriminants based equivalent formula (15) for $QP(\rho)$ based on the corresponding block structure.

The next proposition follows fairly directly from the definition and left to a reader

Proposition C.5. 1. *Consider a block matrix $\rho = \rho(i_1, i_2, j_1, j_2$ as in* (3) (*not necessarily positive semidefinite*). *Additionally to the blocks define also the following N^2-tuple of $N \times N$ matrices*:

$$C_{j_1,j_2} =: \{\rho(i_1, i_2, j_1, j_2) : 1 \leqslant i_1, j_2 \leqslant N\}, \quad 1 \leqslant j_1, j_2 \leqslant N.$$

Associate with the matrix ρ the following two operators

$$T : M(N) \to M(N), T(X) = \sum_{1 \leqslant i,j \leqslant N} X(i,j) A_{i,j},$$

and $T' : M(N) \to M(N)$, $T'(X) = \sum_{1 \leqslant i,j \leqslant N} X(i,j) C_{i,j}$.
Let X be a random complex zero mean matrix such that $E(|X(i,j)|^2 \equiv 1$ and for any permutation $\tau \in S_N$ the random variables $\{X_{i,\tau(i)} : 1 \leqslant i \leqslant N\}$ are independent. Then

$$QP(\rho) = E(det(T(X))\overline{det(X)}) = E(det(T'(X))\overline{det(X)}). \tag{C.3}$$

2. *Consider a $N \times N$ matrix A and a random zero mean vector $z \in C^N$ such that $E(z_i \overline{z_j}) = 0$ for all $i \neq j$. Then*

$$Per(A) = E\left(\prod_{1 \leqslant i \leqslant N} (Az)_i \overline{\prod_{1 \leqslant i \leqslant N} z_i}\right). \tag{C.4}$$

Let us present now the promised short probabilistic proof of (24).

Proposition C.6. *Consider two N^2-tuple of $N \times N$ complex matrices*:

$$\mathbf{A} = (A_{(1,1)}, \ldots, A_{(N,N)}),\ \mathbf{B} = (B_{(1,1)}, \ldots, B_{(N,N)}).$$

Define the following associated four-dimensional tensor:

$$\rho(i_1, i_2, j_1, j_2) =: \sum_{1 \leqslant l \leqslant N^2} A_l(i_2, i_1)\overline{B_l(j_2, j_1)}, \quad 1 \leqslant i_1, i_2, j_1, j_2 \leqslant N. \tag{C.5}$$

Also define two determinantal polynomials:

$$p(X) = det\left(\sum_{1 \leqslant i,j \leqslant N^2} X(i,j) A_{(i,j)}\right), q(X) = det\left(\sum_{1 \leqslant i,j \leqslant N^2} X(i,j) B_{(i,j)}\right).$$

Then $< p, q >_G = < p, q >_{L_2(C^k,\mu)} = QP(\rho)$.

Proof. Consider the following two linear operators $T : M(N) \to M(N)$, $S : M(N) \to M(N)$:

$$T(X) = \sum_{1 \leqslant i,j \leqslant N} X(i,j) A_{(i,j)}, S(X) = \sum_{1 \leqslant i,j \leqslant N} X(i,j) B_{(i,j)}.$$

Then the polynomial $p = det \circ T$ and the polynomial $q = det \circ S$. It follows from Lemma C.2 that

$$< p, q >_{L_2(C^k,\mu)} = < det \circ T, det \circ S >_{L_2(C^k,\mu)} = < det \circ TS^*, det >_{L_2(C^k,\mu)}.$$

Notice that if an $N \times N$ matrix $Y = TS^*(X)$ then

$$Y((i_1, i_2) = \sum_{1 \leqslant j_1, j_2 \leqslant N} \rho(i_1, i_2, j_1, j_2) X(j_1, j_2), 1 \leqslant i_1, i_2 \leqslant N,$$

where ρ is defined in (52). It follows from Eq. (50) that

$$< det \circ TS^*, det >_{L_2(C^k,\mu)} = QP(\rho).$$

Indeed,

$$< det \circ TS^*, det >_{L_2(C^k,\mu)} = E_X(det(TS^*(X))\overline{det(X)}),$$

where the random complex $N \times N$ matrix has a gaussian density $p(X) = \frac{1}{\pi^{N^2}} e^{-(tr(XX^*))}$. That is, the entries $X(i, j)$ are IID canonical complex gaussian random variables, and therefore they are independent, zero mean and $E(|X(i, j)|^2) = 1$, $1 \leqslant i, j \leqslant N$.

(*Formula* (*24*) *is a particular case when* $A_{(i,j)} = (B_{(i,j)}$, $1 \leqslant i, j \leqslant N$.)

Similarly, the permanental formula (43) can be proved using (51). □

Appendix D. Proof of Theorem (5.2)

We assume in this section that $T : M(N) \to M(N)$ is a linear positive (not necessarily completely) operator, the entries of its Choi's representation $CH(T)$ are integer numbers and the maximum magnitude of the entries is **M**. We also assume that

$$T(I) \succ 0, T^*(I) \succ 0. \tag{D.1}$$

Define the following two maps:

$$G(X) = (T(X))^{-1}, F(X) = (T^*(X))^{-1}, \quad X \succ 0.$$

We will also define two discrete dynamical systems. First, the exact one

$$X_{n+1} = G(F(X_n)) = (T((T^*(X_n))^{-1}))^{-1}, X_0 = (T(I))^{-1}. \tag{D.2}$$

Second, the perturbed (i.e. rounded) one

$$Y_{n+1} = G(F(Y_n)) + \delta_n = (T((T^*(Y_n))^{-1}))^{-1} + \delta_n, Y_0 = (T(I))^{-1}, \tag{D.3}$$

where the matrices δ_n, $n \geqslant 0$ are hermitian.

We will need the facts from the next elementary proposition, proof of which is left to the reader.

Proposition D.1. 1. *If* $X \succeq Y$, *i.e.* $X - Y \succeq 0$, *then* $T(X) \succeq T(Y)$ *and* $T^*(X) \succeq T^*(Y)$

2. *Suppose that A is* $N \times N$ *non-singular integer matrix and* $|A(i, j)| \leqslant M$, *the eigenvalues* (*possibly complex*) *of A are* $(\lambda_1, \ldots, \lambda_N)$ *and* $0 < |\lambda_1| \leqslant |\lambda_2| \leqslant \cdots \leqslant |\lambda_N|$. *Then* $|\lambda_N| \leqslant MN$ *and* $|\lambda_1| \geqslant \frac{1}{(MN)^{N-1}}$.

3. *Let* $T : M(N) \to M(N)$ *be a linear positive* (*not necessarily completely*) *operator, the entries of its Choi's representation* $CH(T)$ *are integer numbers and the maximum magnitude of the entries is* **M**. *If inequalities* (53) *are satisfied then*

$$\frac{1}{(\mathbf{M}N^2)^{N-1}} I \preceq T(I), T^*(I) \preceq (\mathbf{M}N^2)I. \tag{D.4}$$

If $\alpha = (\mathbf{M}N^2)^{N-1}$ *then*

$$\alpha^{-1} I \preceq T(I), T^*(I) \preceq \alpha I. \tag{D.5}$$

4. *If* $X \succ 0$ *and* $l(X)$ *is the smallest eigenvalue of* X, $U(X)$ *is the largest eigenvalue of* X *then*

$$l(F(X)), l(G(X)) \geqslant \alpha^{-1} l(X); U(F(X)), U(G(X)) \leqslant \alpha U(X). \tag{D.6}$$

5. *The dynamical system* (54) *is defined for all* $n \geqslant 0$, *i.e. the corresponding matrices involved in* (54) *are strictly positive definite, and thus are non-singular. Moreover,*

$$l(X_{n+1}) \geqslant \alpha^{-2} l(X_n) \text{ and } U(X_{n+1}) \leqslant \alpha^2 U(X_n). \tag{D.7}$$

As $X_0 = (T(I))^{-1}$ *thus from inequalities* (56) *we get that* $l(X_0) \geqslant \frac{1}{\mathbf{M}N^2}$, $U(X_0) \leqslant \alpha$. *Therefore the following inequalities hold*:

$$l(X_n) \geqslant \alpha^{-2(n+1)}, U(X_n) \leqslant \alpha^{2(n+1)}; n \geqslant 0. \tag{D.8}$$

6. *Consider two non-singular square matrices* A *and* B. *Then* $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$ *and thus the following inequality holds*:

$$||A^{-1} - B^{-1}|| \leqslant ||A^{-1}||\, ||A - B||||B^{-1}||, \tag{D.9}$$

where $||A||$ *is, for instance, the largest singular value of* A.

Remark D.2. The main point of Proposition D.1 are inequalities (60): as $\log(\alpha) \leqslant poly(N, \log(\mathbf{M}))$ we get that all matrices involved in (54) have entries of the magnitude at most $2^{poly(N, \log(\mathbf{M}))}$ as long n is at most $poly(N, \log(\mathbf{M}))$.

The next lemma is very essential for our analysis.

Lemma D.3. *Define* $\gamma = \alpha^2$. *If in* (55) *we have that*

$$||\delta_n|| \leqslant \frac{1}{2^{n+1}\gamma^{n+1}} l((T(I))^{-1}), \quad n \geqslant 0,$$

then for all $n \geqslant 0$ *the matrices* $Y_n \succ 0$ *and*

$$l(Y_n) \geqslant \frac{1}{2^n \gamma^n} l((T(I))^{-1}). \tag{D.10}$$

Proof. Consider first the case when $||\delta_n|| \leqslant \frac{1}{2}\frac{l(Y_n)}{\gamma}$, $n \geqslant 0$. Then using the first inequality in (59) we get that

$$l(Y_{n+1}) \geqslant \frac{1}{\gamma} l(Y_n) - \frac{1}{2}\frac{1}{\gamma} l(Y_n) = \frac{1}{2\gamma} l(Y_n).$$

Therefore in this case

$$l(Y_n) \geqslant \frac{1}{2^n \gamma^n} l(Y_0) = \frac{1}{2^n \gamma^n} l((T(I))^{-1}).$$

The general case is done by the induction : for $n = 0$ we can just repeat the argument above. Suppose that inequality (62) holds for all $n \leqslant k$. Then

$$l(Y_{k+1}) \geqslant \frac{1}{\gamma} l(Y_k) - \frac{1}{2^{k+1}\gamma^{k+1}} l((T(I))^{-1}) \geqslant \frac{1}{\gamma}\frac{1}{2^k\gamma^k} l((T(I))^{-1})$$
$$-\frac{1}{2^{k+1}\gamma^{k+1}} l((T(I))^{-1}) = \frac{1}{2^{k+1}\gamma^{k+1}} l((T(I))^{-1}). \qquad \square$$

The next lemma is the punch line.

Lemma D.4. *For any two polynomials $R(N, \log(\mathbf{M}))$ and $S(N, \log(\mathbf{M}))$ there exists a third polynomial $p(N, \log(\mathbf{M}))$ such that if in* (55)

$$||\delta_n|| \leqslant 2^{-p(N,\log(\mathbf{M}))}, \quad n \geqslant 0, \tag{D.11}$$

then

$$||X_n - Y_n|| \leqslant 2^{-S(N,\log(\mathbf{M}))} \ for\ all\ 0 \leqslant n \leqslant R(N, \log(\mathbf{M})). \tag{D.12}$$

Here the matrices $X_n, n \geqslant 0$ satisfy (54)*, the matrices $Y_n, n \geqslant 0$ satisfy* (55)*.*

Proof. Supposed the matrices $\delta_n, n \geqslant 0$ in (55) satisfy inequalities

$$||\delta_n|| \leqslant 2^{-poly(N,\log(\mathbf{M}))} \frac{1}{2^{n+1}\gamma^{n+1}} l((T(I))^{-1}), n \geqslant 0, \tag{D.13}$$

where $poly(N, \log(\mathbf{M}))$ is a sufficiently large polynomial of sufficiently large degree. It follows from Lemma D.3 that, at least, $Y_n \succ 0, n \geqslant 0$. Using (61) we get that

$$||G(F(Y_n)) - G(F(X_n))|| \leqslant ||G(F(Y_n))||||G(F(X_n))||\, ||T(F(Y_n) - F(X_n))||. \tag{D.14}$$

Using (58) we get that

$$||T(F(Y_n) - F(X_n))|| \leqslant \alpha ||F(Y_n) - F(X_n)||. \tag{D.15}$$

Using (61) and (58) again we get that

$$||F(Y_n) - F(X_n)|| \leqslant \alpha ||(T^*(X_n))^{-1}||\, ||(T^*(Y_n))^{-1}||\, ||X_n - Y_n||. \tag{D.16}$$

By the definitions of (54) and (55), we get that

$$||X_{n+1} - Y_{n+1}|| \leqslant ||G(F(Y_n)) - G(F(X_n))|| + ||\delta_n||. \tag{D.17}$$

Denote $||X_n - Y_n|| = Er_n$. Then putting together all these inequalities (66–69) and (62) we finally get that

$$r_{n+1} \leqslant 2^{nRate(N,\log(\mathbf{M}))} r_n + ||\delta_n||, r_0 = 0;\ n \geqslant 0, \tag{D.18}$$

where $Rate(N, \log(\mathbf{M}))$ is some polynomial.
It is clear now that we can satisfy (64) by choosing in (65) a sufficiently large polynomial $poly(N, \log(\mathbf{M}))$ of sufficiently large degree. $\square$

Lemma D.4 is all we need to prove Theorem 5.2 : if matrices δ_k, $k \geqslant 0$ in the statement of Theorem 5.2 satisfy inequality (63), or equivalently inequality (65) for a sufficiently large polynomial $poly(N, \log(\mathbf{M}))$ of sufficiently large degree, then:

1. The matrices

$$P_k, \widetilde{P_k}, Q_k, \widetilde{Q_k}; T(P_k), T(\widetilde{P_k}), T^*(Q_k), T^*(\widetilde{Q_k})$$

   are of at most $2^{H((N,\log(\mathbf{M}))}$ for some universal, i.e. independent from $poly(N, \log(\mathbf{M}))$, polynomial $H((N, \log(\mathbf{M}))$ for $k \leqslant N^2(N\ln(N) + N(\ln(N) + \ln(\mathbf{M}))$.
2. The norms of all "errors" :

$$\|P_k - \widetilde{P_k}\|, \|Q_k - \widetilde{Q_k}\|; \|T(P_k) - T(\widetilde{P_k})\|, \|T^*(Q_k) - T^*(\widetilde{Q_k})\|$$

   are at most $2^{-S(N,\log(\mathbf{M}))}$ for $k \leqslant N^2(N\ln(N) + N(\ln(N) + \ln(\mathbf{M}))$.
3. It remained to choose $poly(N, \log(\mathbf{M}))$ such that $2^{-S(N,\log(\mathbf{M}))}2^{H((N,\log(\mathbf{M}))} \leqslant \frac{1}{N^2}$.
   Then (with a lot of extra room) we get that
   $|\widetilde{DS(T_n)} - DS(T_n)| \leqslant \frac{1}{2N}$ for all $0 \leqslant n \leqslant N^2\,(N\ln(N) + N(\ln(N) + \ln(\mathbf{M})))$.

References

[1] R.B. Bapat, Mixed discriminants of positive semidefinite matrices, Linear Algebra Appl. 126 (1989) 107–124.
[2] A.I. Barvinok, Computing mixed discriminants, mixed volumes, and permanents, Discrete Comput. Geom. 18 (1997) 205–237.
[3] A.I. Barvinok, Polynomial time algorithms to approximate permanents and mixed discriminants within a simply exponential factor, Random Struct. Algorithms 14 (1999) 29–61.
[4] B. Beauzamy, Products of polynomials and a priori estimates for coefficients in polynomial decompositions: a sharp result, J. Symbolic Comput. 13 (1992) 463–472.
[5] B. Beauzamy, E. Bombieri, P. Enflo, H.L. Montgomery, Products of polynomials in many variables, J. Number Theory 36 (1990) 219–245.
[6] A. Ben-Tal, A. Nemirovski, Robust convex optimization, Math. Oper. Res. 23 (4) (1998) 769–805.
[7] L.M. Bregman, A proof of convergence of the method of G.V. Šeleĭhovskiĭfor a problem with transportation-type constraints, Zh. vychisl. Mat. mat. Fiz. 7 (1967) 147–156 (in Russian).
[8] A. Chistov, G. Ivanyos, M. Karpinski, Polynomial time algorithms for modules over finite dimensional algebras, Proceedings of the ISSAC'97, Maui, Hawaii, USA, 1997, pp. 68–74.
[9] M.-D. Choi, Completely positive linear maps on complex matrices, Linear Algebra Appl. 10 (1975) 285–290.
[10] J. Edmonds, System of distinct representatives and linear algebra, J. Res. Nat. Bur. Standards 718 (4) (1967) 242–245.
[11] G.P. Egorychev, The solution of van der Waerden's problem for permanents, Adv. Math. 42 (1981) 299–305.
[12] D.I. Falikman, Proof of the van der Waerden's conjecture on the permanent of a doubly stochastic matrix, Mat. Zametki 29(6) 957. (1981) 931–938 (in Russian).
[13] J. Forster, A linear lower bound on the unbounded error probabilistic communication complexity, Sixteenth Annual IEEE Conference on Computational Complexity, 2001.
[14] M. Grötschel, L. Lovasz, A. Schrijver, Geometric Algorithms and Combinatorial Optimization, Springer, Berlin, 1988.
[15] L. Gurvits, Unbiased non-negative valued random estimator for permanents of complex positive semidefinite matrices, LANL Unclassified Report LAUR 02-5166, 2002.
[16] L. Gurvits, Van der Waerden Conjecture for Mixed Discriminants, arXiv.org preprint math.CO/o406420, 2004.
[17] L. Gurvits, Quantum Matching Theory (with new complexity-theoretic, combinatorial and topological insights on the nature of the Quantum Entanglement), arXiv.org preprint quant-ph/02010222, 2002.

[18] L. Gurvits, Determinantal polynomials, bipartite mixed quantum states, Wick formula and generalized permanental inequalities, 2002, in preparation.
[19] L. Gurvits, Classical deterministic complexity of Edmonds' problem and Quantum Entanglement, arXiv.org preprint quant-ph/0303055, 2003.
[20] L. Gurvits, H. Barnum, Largest separable balls around the maximally mixed bipartite quantum state, Phys. Rev. A 66 (2002) 062311.
[21] L. Gurvits, A. Samorodnitsky, A deterministic polynomial-time algorithm for approximating mised discriminant and mixed volume, in: Proceedings of the 32 ACM Symposium on Theory of Computing, ACM, New York, 2000.
[22] L. Gurvits, A. Samorodnitsky, A deterministic algorithm approximating the mixed discriminant and mixed volume, and a combinatorial corollary, Discrete Comput. Geom. 27 (2002) 531–550.
[23] L. Gurvits, P. Yianilos, The deflation-inflation method for certain semidefinite programming and maximum determinant completion problems, NECI Technical Report, 1998.
[24] V. Kabanets, R. Impagliazzo, Derandomizing polynomial identity tests means proving circuit lower bounds, Electron. Colloq. Comput. Complex 55 (2002).
[25] N. Linial, A. Samorodnitsky, A. Wigderson, A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents, in: Proceedings of the 30th ACM Symposium on Theory of Computing, ACM, New York, 1998.
[26] H. Minc, Permanents, Addison-Wesley, Reading, MA, 1978.
[27] A. Nemirovski, personal communication, 2001.
[28] E. Pascal, Die Determinanten, Teubner-Verlag, Leipzig, 1900.
[29] B. Reznick, An inequality for products of polynomials, Proc. Amer. Math. Soc. 117 (4) (1993) 1063–1073.
[30] G.W. Soules, The rate of convergence of Sinkhorn balancing, Linear Algebra Appl. 150 (1991) 3–40.
[31] L. Valiant, Completeness classes in algebra, in: Proceedings of the 11th ACM Symposium on Theory of Computing, 1979, pp. 249–261.
[32] D.B. Yudin, A.S. Nemirovskii, Informational complexity and efficient methods for the solution of convex extremal problems, Ekonom. i Mat. Metody 12 (1976) 357–369 (in Russian).
[33] D. Zeilberger, Chu's 1303 identity implies Bombieri's 1990 norm-inequality [Via an identity of Beauzamy and Degot], Amer. Math. Monthly (1994).
[34] A. Zvonkin, Matrix integrals and map enumeration: an accessible introduction, Math. Comput. Modelling 26 (8–10) (1997) 281–304.